



Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

QUANTITATIVE IMAGE ANALYSIS FOR THE ASSESSMENT OF PAROTID
GLAND DAMAGE AND TOXICITY PREDICTION FOLLOWING RADIOTHERAPY
TO THE HEAD AND NECK

A THESIS SUBMITTED TO THE UNIVERSITY OF DUBLIN, TRINITY COLLEGE
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY (PhD)

2024

ELIZABETH FORDE

DECLARATION

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish copyright legislation and Trinity College library conditions of use and acknowledgement. I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

A handwritten signature in black ink, appearing to read 'eforde', is written over a horizontal line.

Elizabeth Forde

March 2024

SUMMARY

Patients treated with radiotherapy for head and neck cancer experience debilitating xerostomia and sticky saliva, caused by damage to the parotid glands. Currently predicting this toxicity is limited by the inability of existing normal tissue complication probability models to account for the complexity of the glands' patient specific anatomy and their heterogeneous biology. Applying quantitative imaging analysis via "radiomics" enables medical images to be converted to minable data which reflects underlying biology of specific regions of interest. These radiomic features could be informative of how normal tissues respond to treatment by documenting patient specific texture of the parotid glands. Hypothetically, radiation-induced xerostomia and sticky saliva could be predicted prior to the manifestation of clinical signs, using radiomic analysis of routine computer tomography (CT) images from patients with head and neck cancer.

Prior to testing this hypothesis, this thesis first investigated specific nuances within the radiomics pipeline which could influence radiomic feature values. Specifically, Study 1 quantified the difference in radiomic feature values that can occur as a result of inter-observer delineation variability of the right parotid gland. Based on a cohort of 40 independent observers, results of this study identified 14 specific radiomic features which are robust to these uncertainties.

It is understood that radiomic features are also influenced by CT scanner acquisition and reconstruction protocols. This variability can be assuaged by pre-processing methods applied prior to feature extraction, for example via spatial resampling and intensity discretisation; however, a knowledge gap existed for this approach when studying the parotid glands. Study 2 therefore explored the impact of resampling to a common voxel size on radiomic feature values. This process was tested across two different bin widths to assess the reproducibility under different discretisation conditions. Based on a cohort of 127 patients, results of this study,

identified several radiomic features which are robust to these pre-processing settings. Results here indicated the combination of isotropic resampling and the use of a smaller bin width of 5 HU may be optimal for radiomic analysis of the parotid glands.

The work conducted in Studies 1 and 2 was a required step for the responsible exploration of the radiomics pipeline; however, they lacked an application to clinical outcomes. This clinical link was subsequently addressed in Studies 3 and 4. In Study 3 logistic regression models were built using a combination of clinical, dosimetric, and radiomic features to predict acute xerostomia and sticky saliva. This study included a cohort of 98 patients treated with chemoradiation for squamous cell carcinoma of the head and neck. The integration of radiomics into the models improved the prediction of acute xerostomia and sticky saliva. Radiation dose is accumulated during a course of treatment, and the subsequent tissue response is therefore dynamic. Based on images acquired at three timepoints in a cohort of 26 patients, Study 4 identified the variability of specific radiomic features that were associated with acute and late xerostomia. Quantifying variation in radiomic features prior to the onset of clinical symptoms in such a way, allows the opportunity for prompt intervention, which may include adaptive radiotherapy.

The work presented in this thesis has demonstrated the viability of radiomic analysis of parotid gland damage following radiotherapy for head and neck cancer. The outcome of this work will ultimately assist in the personalisation of radiotherapy and supportive care for our patients.

LAY SUMMARY

Patients with head and neck cancer experience significant side effects from treatment; for example, a dry mouth and saliva that becomes sticky. This is mainly due to the damage of salivary glands which occurs during radiotherapy. Identifying which patients are at a higher risk of experiencing these side effects is challenging. Predicting treatment related side effects is primarily based on estimates of the amount of radiation the salivary glands receive during treatment; however, this approach does not consider patient specific characteristics.

Medical imaging allows for the visualisation of gross anatomy, but may also provide the opportunity to reveal underlying biology of individual patient's salivary glands. This added detail could be used to supplement existing clinical data to better predict how salivary glands respond to radiation. This novel approach to analysing images, is technically challenging and can be influenced by several factors, such as how the image is taken or how the anatomical region is outlined. The work presented in this thesis explored the impact of these factors on the data generated from analysing the medical images, as well as whether this data could be used to predict side effects related to damage of the salivary glands.

In essence, this thesis has analysed numerical data taken from medical imaging of the salivary glands to better predict the incidence of dry mouth and sticky saliva following radiotherapy for head and neck cancer.

LIST OF ABBREVIATIONS

3D	3 Dimensions
3DCRT	3 Dimensional Conformal Radiotherapy
5-FU	5-Flurouracil
AUC-ROC	Area Under the Curve Receiver Operator Curve
CCR	Credence Cartridge Radiomics
CI	Confidence Interval
CIP	Cancer Imaging Program
COM	Centre of Mass Shift
COV	Coefficient of Variance
CRT	Chemoradiotherapy
CT	Computer Tomography
CTCAE	Common Terminology for Adverse Events
DICOM	Digital Imaging and Communications in Medicine
D_{mean}	Mean Dose
DPIA	Data Protection Impact Assessment
DSC	Dice Similarity Coefficient
DVC	Dose Volume Constraint
DVH	Dose Volume Histogram

ECMP	European Congress of Medical Physics
EGFR	Epidermal Growth Factor Receptor
EHNS	European Head and Neck Society
EORTC	European Organisation for Research and Treatment of Cancer
ESMO	European Society for Medical Oncology
ESTRO	European Society of Radiotherapy and Oncology
EU	European Union
FALCON	Fellowship in Anatomic DeLineation and CONtouring
FDG	Fludeoxyglucose
FOV	Field of View
GDPR	General Data Protection Regulation
gEUD	Generalised Equivalent Uniform Dose
GLCM	Grey level Co-occurrence Matrix
GLRLM	Grey-level Run Length Matrix
GLRLM_GLNU	Grey Level Non Uniformity
GLRLM_HGRE	High Grey Level Run Emphasis
GLRLM_LGRE	Low Grey Level Run Emphasis
GLRLM_LRE	Long Runs Emphasis
GLRLM_LRHGE	Long Run High Grey Level Emphasis
GLRLM_LRLGE	Long Run Low Grey Level Emphasis

GLRLM_RLNU	Run Length Non Uniformity
GLRLM_RP	Run Percentage
GLRLM_SRE	Short Runs Emphasis
GLRLM_SRHGE	Short Run High Grey Level Emphasis
GLRLM_SRLGE	Short Run Low Grey Level Emphasis
GLSZM	Grey-level Zone Length Matrix
GLZLM_GLNU	Grey Level Non Uniformity
GLZLM_GLV	Grey Level Variance
GLZLM_HGZE	High Gray Level Zone Emphasis
GLZLM_LGZE	Low Gray Level Zone Emphasis
GLZLM_LZE	Large Zone Emphasis
GLZLM_LZHGE	Large Zone High Grey Level Emphasis
GLZLM_LZLGE	Large Zone Low Grey Level Emphasis
GLZLM_NGLNU	Normalised Grey Level Non Uniformity
GLZLM_NZLNU	Normalised Zone Size Non Uniformity
GLZLM_SZE	Small Zone Emphasis
GLZLM_SZHGE	Small Zone High Grey Level Emphasis
GLZLM_SZLGE	Small Zone Low Grey Level Emphasis
GLZLM_ZLNU	Zone Size Non Uniformity
GLZLM_ZP	Zone Percentage
GLZLM_ZSE	Zone Size Entropy
GLZLM_ZSV	Zone Size Variance

GTV_T	Gross Tumour Volume_Tumour
HBOT	Hyperbaric Oxygen Therapy
HD	Hausdorff Distance
HNC	Head and Neck Cancer
HPV	Human Papillomavirus
HU	Hounsfield Unit
IBEX	Imaging Biomarker Explorer
IBSI	Image Biomarker Standardisation Initiative
ICC	Intra Class Coefficient
IMRT	Intensity Modulated Radiation Therapy
IQR	Inter Quartile Range
LASSO	Least Absolute Shrinkage and Selection Operator
LIFEx	Local Image Feature Extraction
LKB	Lyman Kutcher Burman
MRI	Magnetic Resonance Imaging
MU	Monitor Units
MV	Mega Voltage
MVCT	Mega Voltage Computer Tomography
NGTDM	Neighbourhood Grey-Level Different Matrix
NIH	National Institute of Health
NMI	Nuclear Medicine Imaging

NTCP	Normal Tissue Complication Probability
OPC	Oropharyngeal Cancer
OR	Odds Ratio
PC	Principal Component
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PROM	Patient Reported Outcome Measures
QLQ	Quality of Life Questionnaire
QLQ-HN35	Quality of Life Questionnaire - Head and Neck
QoL	Quality of Life
QUANTEC	Quantitative Analysis of Normal Tissue Effects in the Clinic
RCT	Randomised Control Trial
REACH	Randomised Trial of Avelumab-cetuximab-radiotherapy versus Standard of Care in Locally Advanced Squamous Cell Carcinoma of the Head and Neck
RF	Radiomic Feature
ROI	Region of Interest
RQS	Radiomics Quality Score
RT	Radiation Therapy
RTOG	Radiation Therapy Oncology Group
RTSTRUCT	Radiotherapy Structure Set

SEER	Surveillance Epidemiology End Results
SOC	Standard of Care
SOMREC	School of Medicine Research Ethics Committee
SPECT	Single Proton Emission Computer Tomography
SPSS	Statistical Package for Social Sciences
SS	Sticky Saliva
SVR	Surface to Volume Ratio
T-stage	Tumour Stage
TCD	Trinity College Dublin
TCGA-HNSC	The Cancer Genome Atlas Head and Neck Squamous Cell Carcinoma
TCIA	The Cancer Imaging Archive
TENS	Transcutaneous Electrical Nerve Stimulation
TNM	Tumour Node Metastasis
TORS	Trans Oral Robotic Surgery
US	Ultrasound
VIF	Variance Inflation Factor
VMAT	Volumetric Modulated Arc Therapy
WHO	World Health Organisation

RESEARCH OUTPUTS

Study 1

1. Forde E, Leech M, Robert C, Herron E, Marignol L. Influence of Inter-Observer Delineation Variability on Radiomics Features of the Parotid Gland. European Society of Radiotherapy and Oncology (ESTRO) Annual Congress Milan, Italy. November 2020.
2. Forde E, Leech M, Robert C, Herron E, Marignol L. Influence of interobserver delineation variability on radiomics features of the parotid gland. *Physica Medica*. 2021; 82: 240-248

Study 2

1. Forde E, Robert C, Leech M, Marignol L. Spatial Resampling Combined with Intensity Discretization Improves Reproducibility of Radiomic Features of the Parotid Gland. European Congress in Medical Physics (ECMP) Annual Meeting, Dublin, Ireland. August 2022.

Study 3

1. Forde E, Leech M, Robert C, Çiflik V, Marignol L. Unveiling Hidden Signals: Radiomics Enhances Prediction of Acute Xerostomia and Sticky Saliva. European Society of Radiotherapy and Oncology (ESTRO) Annual Congress, Glasgow, Scotland. May 2024.
2. Forde E, Leech M, Marignol L. Unlocking Precision in Radiotherapy: Radiomic Analysis for Predicting Radiation-Induced Xerostomia and Sticky Saliva. Trinity St James Cancer Institute Research Blitz. Dublin, Ireland. December 2023.

Study 4

1. Forde E, Leech M, Marignol L. Unveiling the Hidden: Radiomic Insights into Parotid Gland Transformations During Radiotherapy for Xerostomia Prediction. The Faculty of Radiologists and Radiation Oncologists Annual Meeting, Dublin Ireland. September 2023.

Invited Talks Related to this Thesis Topic

1. Forde, E. Management of Radiation Toxicity in Head and Neck Cancers. European Society of Radiotherapy and Oncology (ESTRO) Annual Congress. Online. May 2020.
2. Forde E. Radiomics in Head and Neck Cancers. European Society of Radiotherapy and Oncology (ESTRO) Annual Congress, Copenhagen, Denmark. May 2022.

ACKNOWLEDGEMENTS

I first presented the initial research proposal for this work at the ESTRO Physics Masterclass for Beginners in Research, in Florence in 2017. I am grateful for this early constructive criticism and support I received from both the faculty and participants on the course. I would also like to express my gratitude to Claudio Fiorino and Martina Mori from San Raffaele Cancer Institute in Milan, where I spent a week discussing in detail their previous experience in radiomics, specifically their work in radiation induced changes to the parotid glands. This trip was funded by the ESTRO Mobility Grant I was awarded in 2018, and the team at San Raffaele were all so generous in their time when I was with them.

I would like to extend my gratitude to Charlotte Robert from the Department of Medical Physics at Gustave Roussy. Charlotte has both encouraged and challenged me over the years, and I value all I have learnt from her. I would also like to acknowledge Christophe Nioche. Although I have never actually met Christophe, he is the name behind the LIFEx customer support email, and has been a frequent contact of mine since I started this research. Both Charlotte and Christophe have been my “LIFEx experts”, and have always been there to answer the most specific of radiomics questions.

With respect to anonymisation and transfer of DICOM data I would like to kindly acknowledge the ESTRO FALCON group, specifically Miika Palmu for arranging this for Study 1. For Study 3, I greatly appreciate the efforts of Vjona Cifliku and Killian Sambourg from Gustave Roussy for the hours of work they put into facilitating the data transfer.

On a daily basis I am thankful for the constant support and inspiration from all my colleagues in the Discipline of Radiation Therapy here at Trinity. This wonderful group of colleagues and friends has kept me calm in times of stress and has guided me to be the researcher I aspire to become. In particular, I am beyond grateful to Laure Marignol and Michelle Leech, my

PhD supervisors, for their guidance and attention to detail in the preparation and presentation of this significant body of work.

Finally, my beautiful family. My husband, Aaron and our two children, Liam and Eimear. After writing this thesis, it seems strange to struggle putting words down on a page, but I just can't express enough what their support means to me. Every single moment shared and memory created has kept me going.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	28
1.1 INCIDENCE AND MANAGEMENT OF HEAD AND NECK CANCER	28
1.2 RADIATION INDUCED XEROSTOMIA AND STICKY SALIVA	38
1.3 MODELLING NORMAL TISSUE DAMAGE	53
1.4 RADIOMICS	56
1.5 CENTRAL HYPOTHESIS AND AIMS OF THIS THESIS	64
CHAPTER 2: METHODOLOGY	65
2.1 PATIENT COHORTS AND RESEARCH ETHICS	65
2.2 RADIOMICS PROCEDURES	67
2.3 STUDY OUTCOMES	81
2.4 STATISTICAL ANALYSES	82
CHAPTER 3 STUDY 1: INFLUENCE OF INTER-OBSERVER DELINEATION VARIABILITY ON RADIOMICS FEATURES OF THE PAROTID GLAND	86
3.1 INTRODUCTION	86
3.2 MATERIALS AND METHODS	87
3.3 RESULTS	90
3.4 DISCUSSION	103
3.5 CONCLUSION	109
CHAPTER 4 STUDY 2: INFLUENCE OF SPATIAL RESAMPLING AND INTENSITY DISCRETIZATION ON THE REPRODUCIBILITY OF RADIOMIC FEATURES OF THE PAROTID GLAND	110
4.1 INTRODUCTION	110
4.2 MATERIALS AND METHODS	111
4.3 RESULTS	114
4.4 DISCUSSION	139
4.5 CONCLUSION	149

**CHAPTER 5 STUDY 3: INTEGRATION OF RADIOMICS INTO PREDICTIVE
MODELLING OF ACUTE XEROSTOMIA AND STICKY SALIVA** **150**

5.1	INTRODUCTION	150
5.2	MATERIALS AND METHODS	151
5.3	RESULTS	183
5.4	DISCUSSION	222
5.5	CONCLUSION	234

**CHAPTER 6 STUDY 4: RADIOMIC INSIGHTS INTO PAROTID GLAND
TRANSFORMATIONS DURING RADIOTHERAPY FOR XEROSTOMIA
PREDICTION** **235**

6.1	INTRODUCTION	235
6.2	MATERIALS AND METHODS	235
6.3	RESULTS	242
6.4	DISCUSSION	291
6.5	CONCLUSION	299

CHAPTER 7: DISCUSSION **300**

7.1	ADVANCING RADIATION TOXICITY PREDICTION THROUGH RADIOMICS	300
7.2	THE POTENTIAL OF RADIOMICS IN ADAPTIVE RADIOTHERAPY	310
7.3	INTEGRATING IMAGING BIOMARKERS WITHIN A CLINICAL AND SOCIAL CONTEXT	315
7.4	RECOMMENDATIONS FOR FUTURE WORK	320

CHAPTER 8: CONCLUSION **325**

CHAPTER 9: REFERENCES **326**

LIST OF FIGURES

- Figure 1.1: Schematic diagram of the cellular composition of salivary gland.
- Figure 1.2: Classes of radiomic features.
- Figure 2.1: Radiomics workflow.
- Figure 3.1: Contours of the right parotid gland defined by the expert, auto-segmentation and three observers.
- Figure 3.2: Plot of the mean absolute paired percentage difference of the participants values from the expert values ranked according to increasing mean.
- Figure 3.3: Scatterplot of the absolute paired percentage difference in feature values between participant and expert against their corresponding DICE similarity coefficient and Hausdorff Distance.
- Figure 3.4: Scatterplot of the least and most robust features, based on the mean absolute paired percentage difference.
- Figure 4.1: Process of methodology for including intensity discretisation, spatial resampling, feature extraction, data analysis, and classification.
- Figure 4.2: Variability of radiomic feature values following spatial resampling expressed as percentage coefficient of variance. Data is based on a fixed bin width of 10 HU.
- Figure 4.3: Variability of radiomic feature values following spatial resampling ranked according to percentage coefficient of variance. Data is based on a fixed bin width of 10 HU.

- Figure 4.4: Variability of radiomic feature values following spatial resampling expressed as percentage coefficient of variance. Data is based on a fixed bin width of 5 HU.
- Figure 4.5: Variability of radiomic feature values following spatial resampling ranked according to percentage coefficient of variance. Data is based on a fixed bin width of 5 HU.
- Figure 4.6: Variability of radiomic feature values following spatial resampling expressed as percentage coefficient of variance. Data are displayed based on a fixed bin width of 5 HU and 10 HU.
- Figure 5.1: Spearman correlation matrix assessing collinearity of radiomic features extracted based on a fixed bin width of 5 HU.
- Figure 5.2: Spearman correlation matrix assessing collinearity of radiomic features extracted based on a fixed bin width of 10 HU.
- Figure 5.3: Scree plot analysis of principal component analysis for radiomic features extracted based on a fixed bin width of 5 HU.
- Figure 5.4: Scree plot analysis of principal component analysis for radiomic features extracted based on a fixed bin width of 10 HU.
- Figure 5.5: Shortened version of CLEAR checklist (CLEAR-S) containing methodologic quality items of Study 3
- Figure 5.6: Trial design chart depicting randomisation of patients enrolled on the EORTC-REACH clinical trial.

- Figure 5.7: Incidence of baseline, acute, sub-acute, and late xerostomia, and sticky saliva
- Figure 5.8: Scatterplot of patient smoking status and baseline xerostomia.
- Figure 5.9: Scatterplot of patient smoking status and baseline sticky saliva.
- Figure 6.1: Incidence of xerostomia reported at mid-treatment, near the end of treatment, and 3-6 months following treatment.
- Figure 6.2: Mean absolute paired percentage difference of all radiomic features per patient.
- Figure 6.3: Box and Whisker plots of features in which there was a statistically significant difference in the absolute paired percentage difference from CT1 to CT2 for the patients who experienced xerostomia.
- Figure 6.4: Box and Whisker plots of features in which there was a statistically significant difference in the absolute paired percentage difference from CT1 to CT2 for the patients who experienced late xerostomia compared to those who did not.
- Figure 6.5: Box and Whisker plots of features in which there was a statistically significant difference in the absolute paired percentage difference from CT2 to CT3 for the patients who experienced late xerostomia compared to those who did not.
- Figure 6.6: Box and Whisker plots of features in which there was a statistically significant difference in the absolute paired percentage difference from CT1 to CT3 for the patients who experienced late xerostomia compared to those who did not.

Figure 6.7: Paired percentage difference of Shape_Volume (mL) per patient measured from CT imaging taken before the start of treatment and at mid-way during treatment (A), mid-way through treatment and at the completion of treatment (B), and before the start of treatment and at the completion of treatment (C).

Figure 6.8: Paired percentage difference of Shape_Sphericity per patient measured from CT imaging taken before the start of treatment and at mid-way during treatment (A), mid-way through treatment and at the completion of treatment (B), and before the start of treatment and at the completion of treatment (C).

Figure 6.9: Paired percentage difference of Shape_Surface per patient measured from CT imaging taken before the start of treatment and at mid-way during treatment (A), mid-way through treatment and at the completion of treatment (B), and before the start of treatment and at the completion of treatment (C).

Figure 6.10: Paired percentage difference of Shape_Compacity per patient measured from CT imaging taken before the start of treatment and at mid-way during treatment (A), mid-way through treatment and at the completion of treatment (B), and before the start of treatment and at the completion of treatment (C).

Figure 7.1: Proposed workflow for radiomics driven adaptive radiotherapy.

LIST OF TABLES

Table 1.1:	Summary of key evidence comparing treatment management options for patients with head and neck cancer.
Table 1.2:	Summary of the pharmacological and non-pharmacological interventions that have been explored in the prevention and management of parotid toxicities.
Table 2.1:	Shape/morphological and intensity histogram radiomic features extracted per study based on the specific version of LIFEx used.
Table 2.2:	Texture features extracted per study based on the specific version of LIFEx used.
Table 2.3:	Summary of radiomics pre-processing and feature extraction process for each study.
Table 2.4:	Summary of data source, patient cohort, and data analysis for each study.
Table 3.1:	Absolute paired percentage difference of each radiomic feature, along with the standard deviation and stability classification.
Table 3.2:	Mean absolute paired percentage difference of each radiomic feature, along with the standard deviation and stability classification.
Table 4.1:	Percentage coefficient of variance, along with classification of reproducibility of radiomic features following spatial resampling of Cohort 1 based on fixed bin widths of 5 HU and 10 HU.

Table 4.2:	Percentage coefficient of variance, along with classification of reproducibility of radiomic features following spatial resampling of Cohort 1 and Cohort 2 based on a fixed bin width of 5 HU.
Table 5.1:	Metadata from the planning CT scans of the 98 patients extracted using OsiriX Lite (version 13.0.2).
Table 5.2:	Comparison of radiomic feature values between the ipsilateral and contralateral parotid glands based on Wilcoxon Signed Rank test and a fixed bin width of 5 HU.
Table 5.3:	Comparison of radiomic feature values between the ipsilateral and contralateral parotid glands based on Wilcoxon Signed Rank test and a fixed bin width of 10 HU.
Table 5.4:	Nine principal components identified with Eigenvalues <1 based on a fixed bin width of 5 HU.
Table 5.5:	Nine principal components identified with Eigenvalues <1 based on a fixed bin width of 10 HU.
Table 5.6:	Collinearity assessment based on Variance inflation factors results following feature reduction, based on a fixed bin width of 5 HU.
Table 5.7:	Collinearity assessment based on Variance inflation factors results following feature reduction, based on a fixed bin width of 10 HU.
Tables 5.8:	Patient and disease characteristics for the full cohort of 98 patients.
Table 5.9:	Crosstabulation of HPV status by history of smoking and alcohol consumption for patients diagnosed with oropharyngeal cancer.

Table 5.10:	REACH trial randomisation of patients amongst trial arms for the full cohort of 98 patients.
Table 5.11:	Results of chi-Squared analysis testing the association between categorical clinical variables and baseline, acute, sub-acute, and late xerostomia/sticky saliva.
Table 5.12:	Results of Mann Whitney U test examining the association between continuous clinical/dosimetric variables and acute, sub-acute, and late xerostomia/sticky saliva.
Table 5.13:	Results of chi-Squared analysis testing the association between baseline xerostomia/sticky saliva and acute, sub-acute, and late xerostomia/sticky saliva.
Table 5.14:	Results of Mann Whitney U test examining the association between each radiomic feature and acute, sub-acute, and late xerostomia/sticky saliva. Results are based on a fixed bin width of 5 HU.
Table 5.15:	Results of Mann Whitney U test examining the association between each radiomic feature and acute, sub-acute, and late xerostomia/sticky saliva. Results are based on a fixed bin width of 10 HU.
Table 5.16:	Logistic regression result for the clinical, radiomic and combined models for the prediction of moderate/severe acute xerostomia based on a fixed bin width of 5 HU.
Table 5.17:	Metrics quantifying each model's performance in predicting moderate/severe acute xerostomia. Models are based on a fixed bin width of 5 HU.
Table 5.18:	Metrics quantifying each model's performance in predicting moderate/severe acute xerostomia. Models are based on a fixed bin width of 5 HU.

Table 5.19:	Logistic regression result for the clinical, radiomic and combined models for the prediction of moderate/severe acute xerostomia based on a fixed bin width of 10 HU.
Table 5.20:	Logistic regression result for the clinical, radiomic and combined models for the prediction of moderate/severe acute sticky saliva based on a fixed bin width of 5 HU.
Table 5.21:	Logistic regression result for the clinical, radiomic and combined models for the prediction of moderate/severe acute sticky saliva based on a fixed bin width of 10 HU.
Table 6.1:	Patient characteristics, treatment details, and RTOG xerostomia scores.
Table 6.2:	RTOG xerostomia scores recorded at mid-treatment, near-end treatment, and 3-6 months post treatment.
Table 6.3:	Friedman test results for radiomic feature values pre-treatment, mid-treatment, and post-treatment.
Table 6.4:	Results of statistical analysis (Wilcoxon signed rank test and effect size) on changes in feature values between scans acquired pre-treatment, mid-treatment, and post-treatment.
Table 6.5:	Descriptive statistics of radiomic features extracted based on pre-treatment, mid-treatment, and post-treatment scans.
Table 6.6:	Comparison of patients with and without acute xerostomia (measured at mid-treatment) based on the absolute paired percentage difference of all patients for each radiomic feature.
Table 6.7:	Comparison of patients with and without acute xerostomia (measured near the end of treatment) based on the

absolute paired percentage difference of all patients for each radiomic feature.

Table 6.8: Comparison of patients with and without late xerostomia (measured 3-6 months following treatment) based on the absolute paired percentage difference of all patients for each radiomic feature.

1.1 INCIDENCE AND MANAGEMENT OF HEAD AND NECK CANCER

1.1.1 Incidence and Epidemiology of Head and Neck Cancer

Head and neck cancer (HNC) is currently the seventh most common cancer globally, with an expected 30% increase in incidence annually by 2030 (1, 2). Based on the International Classification of Diseases from the World Health Organisation (WHO), the term “head and neck cancer” includes 19 anatomical sub-sites within the region, 90% of which are squamous cell in origin (3). Historically, HNC has predominately been associated with tobacco use and alcohol consumption; behaviours which are thought to account for a >35-fold increased risk of developing these cancers (2). Whilst these risks remain, the differing aetiology of anatomical subsites within the head and neck region have prevailed in recent years, largely attributed to the impact of viral exposure; for example, Epstein Barr virus in the case of nasopharyngeal cancer and Human Papillomavirus (HPV) in the case of oropharyngeal cancer (OPC) (4, 5).

Despite the availability of the HPV vaccine, we are still seeing an increase in HPV associated OPCs today, and it is predicted that over the next 20 years, OPC will overtake oral cavity cancers in many European countries (3, 4, 6). This may be due to the interval time between HPV exposure and the development of OPC, which is likely to exceed 10 years (7). It has also been established that patients with HPV positive tumours are more likely to present with smaller primary tumours and more extensive nodal disease (8). Furthermore, it is now understood that HPV positive HNCs are more radiosensitive due to there being fewer hypoxic regions within the tumour microenvironment (9). This biological understanding, combined with improvements in radiotherapy technologies and improved understanding

of the synergistic effects of adjuvant chemotherapy has led to changes in management and an improvement in survival outcomes for this ever-growing patient group. Given OPC incidence is increasing at the fastest rate (3), effective management of this specific HNC subtype is paramount.

1.1.2 Management of Oropharyngeal Cancer

The inclusion of radiotherapy in the management of OPC has evolved over the years, making it the standard of care for both early and advanced stage disease. For early-stage disease, i.e., T1-2N0M0, irrespective of HPV status, single modality treatment is preferred, including either surgical resection in the form of transoral robotic surgery (TORS) or radiation therapy (RT) using image guided modulated techniques (10). Retrospective and non-randomised evidence suggests equivalent oncological outcomes with comparable rates of overall survival and local regional control between the two approaches (11, 12). This was confirmed in a meta-analysis based on level 4 evidence of over 1000 patients which indicated a 5-year disease specific survival of 90.4% and 89.6% for RT and TORS respectively (13). Whilst TORS results in greater organ preservation compared to traditional open surgical techniques, the extent of the resection may still result in functional morbidity, for example dysphagia (14). This raises the question whether RT would be preferable given it is less invasive; however, even in the case of early-stage disease, post-operative RT may still be indicated due to anatomical restrictions impacting on adequate access to the site, or due to pathological upstaging in the case of positive microscopic margins (R1 resection). (10). Here, we see the advantages of RT, in particular modulated techniques which are able to deliver highly conformal doses, thereby “accessing” the site of primary disease whilst limiting treatment related morbidity.

In the case of locally advanced stage disease (T3-4, N2-N3) or disease with high-risk features (i.e., positive margins, extracapsular nodal spread, perineural or lymphovascular invasion) adjuvant RT with or without chemotherapy is also required to optimise local control (10, 15, 16). The

multimodality approach of post-operative chemoradiation is supported by RCT data indicating a survival benefit for high-risk patients over post-operative RT alone (17, 18). Although surgical morbidity with the use of TORS remains low, the prospect of using chemoradiation in a definitive manner has also been explored. This approach is now considered the standard of care for this locally advanced stage disease, irrespective of HPV status.

With respect to radiotherapy doses, when using conventional fractionation, the current standard of care for definitive treatment of locally advanced stage disease is to treat the primary lesion to a total dose of 70 Gy, along with the high-risk disease/lymph nodes and elective nodes to 63 Gy and 50 Gy respectively (19), although doses may vary (20). When clinically appropriate, the use of hyperfractionated schedules treating 1.2Gy per fraction, delivered bi-daily to a total dose of 81.6Gy may also be used; however, consensus is lacking regarding the optimal altered fractionation schedules, and the potential increase in toxicity should be considered (19). In the case of adjuvant treatment, the dose to the tumour bed may be reduced to 60 Gy, although if positive surgical margins are observed it is recommended to increase this dose to the region of 63-66 Gy (15, 21). In these cases, the surgical bed and subclinical disease should be simultaneously boosted to doses of 50 Gy (15). Whether CRT is delivered definitively or adjuvantly, dose to the primary disease may be delivered with conventional or moderately hypofractionated schedules (i.e., 2 Gy – 2.2 Gy per fraction) (21). Although the meta-analysis conducted by Jiang et al., (22) failed to demonstrate a clinical advantage of the simultaneous integrated boost technique over the traditional sequential boost approach, the former is often preferred due to its convenience. As such, all other targets are typically treated simultaneously over the same 35 fractions which, when treating once daily, results in an overall treatment duration of 7 weeks (23). Given the risk of accelerated repopulation that can occur with squamous cell carcinoma, treatment may also be delivered 6 times per week, reducing the overall treatment schedule to 6 weeks. (24).

Based on an updated meta-analysis of 107 trials with nearly 20 000 patients, the use of concomitant chemotherapy with radiation has shown superiority compared to radiotherapy alone with an absolute benefit of 6.5% and 3.6% at 5 and 10 years respectively (25). In light of this evidence, for patients eligible to receive platinum-based agents, 100mg/m² cisplatin administered once every three weeks is recommended in the EHNS-ESMO-ESTRO clinical practice guidelines, irrespective of HPV status (10). Although more toxic than if 30mg/m² is delivered once per week instead, the 2-year local regional control rate was increased by nearly 15% with the 3-weekly treatment (26). Acute adverse events which were statistically significant ($p < 0.05$) between the two arms in this RCT included vomiting, infection, deafness, hyponatremia, leukopenia, neutropenia, and lymphocytopenia. Alternative, viable options to high-dose, single agent cisplatin supported by randomised data include either 5-fluorouracil (5-FU) plus carboplatin, or carboplatin plus paclitaxel (15). Despite these numerous options, not all patients will be fit to receive platinum-based chemotherapy due to impaired renal function (27). For these patients, there is scope to intensify treatment not through altered radiotherapy fractionation schedules or specific chemotherapy agents, but via the inclusion of monoclonal antibodies and immune checkpoint inhibitors.

Cetuximab is one example of a monoclonal antibody that inhibits epidermal growth factor receptor (EGFR) which mediates tumour cell growth. When combined with chemotherapy, this action results in “immunogenic cell death” (28). Although recommended by the EHNS-ESMO-ESTRO clinical practice guidelines for selected patients, the use of weekly concomitant cetuximab has not been shown to improve progression free survival and overall survival, even in the prognostically favourable HPV positive patients (10, 27, 29). To complement cetuximab, the addition of avelumab (a PD-1/PD-L1 checkpoint inhibitor) may be beneficial. This monoclonal antibody selectively binds to PD-L1 on tumour cells, blocking their interaction with PD-1 T-cells, ultimately limiting tumour growth potential (30, 31). When used in combination, cetuximab and avelumab could work synergistically to amplify the immune response in the tumour microenvironment (28). The

safety and efficacy of this combination has been recently tested in the EORTC RCT: Avelumab – cetuximab - radiotherapy versus standards of care in locally advanced squamous-cell carcinoma of the head and neck (REACH) (32). The preliminary safety phase of this trial, based on 82 patients, demonstrated acceptable levels of adverse events. Specifically, the grade 3 toxicities reported in the experimental arms were mucositis, dysphagia, and radiodermatitis; the incidence of which was consistent with the standard of care previously reported (30).

Patients with HPV positive OPC are likely to respond more favourably compared to patients with HPV negative disease due to the increased radiosensitivity and difference in the clinicopathological characteristics of this tumour subtype (33). Not only has this questioned the need for the use of radiosensitisers, such as Nimorazole, for this specific population (34, 35), but this favourable response profile has also questioned the need for such aggressive treatment regimens (36). Despite the promising results of initial studies, treatment de-intensification is currently non-standard practice, owing to the lack phase 3 RCTs demonstrating an acceptable survival benefit (36). As such, the current SOC, which includes the combination of high dose radiotherapy, cytotoxic agents, with or without monoclonal antibodies, whilst effective, does result in significant acute and late toxicity. The impact of these toxicities is compounded considering the emerging profile of patients with OPC. The increase in incidence of HPV positive oropharyngeal cancers means we are now treating younger patients with a better prognosis. As such, greater attention must be paid to treatment related toxicities that impact quality of life during the ever-extending survivorship period. The above key literature discussed is summarised in Table 1.1.

Author, Year	Study Design	Study Aim	Participants	Context (Intervention/ Comparison)	Primary Endpoint	Main Finding
Pedro et al., 2018	Retrospective cohort study	Review outcomes of early-stage OPSCC treated with surgery or primary RT	OPSCC T1-T2 cN0-N1 (n= 61)	Comparison between surgery (with or without adjuvant RT) versus primary RT (with or without chemotherapy).	3- and 5-year OS and LRFS	No significant difference in OS and LRFS between both treatments
Morisod et al., 2016	Meta-analysis based of Level 4 evidence	Compare survival outcomes between transoral surgery and RT for early stage oropharyngeal SCC.	OPSCC T1–2, N0, M0 (n= 1005; RT group = 729, Surgical group =276 patients).	Transoral Surgery vs RT	2-year, 3-year, or 5-year DSS	Equivalent efficacy of treatment techniques
Bernier et al., 2004	Randomized controlled trial	To compare concomitant cisplatin and RT vs RT alone LA-HNC	Stage III or IV HNC treated post-operatively (n=334).	Concurrent CRT vs RT alone	PFS	Statistically significant difference in progression-free survival, favouring the concurrent CRT over RT alone
Cooper et al., 2004	Randomized controlled trial	To determine whether concurrent cisplatin therapy and postoperative radiotherapy improve	High-risk HNSCC treated post-operatively (n=459)	RT alone vs CRT	LRC	Concurrent postoperative CRT significantly improves local and regional control and disease-free survival; albeit with increased toxicity

the rates of local and regional control among patients						
Overgaard et al., 2010	Randomised, open-label trial	To determine whether accelerated fractionation could be applied in developing countries for HNSCC treatment	Patients with HNSCC of the larynx, pharynx, and oral cavity eligible for curative radiotherapy (n = 908 total)	Accelerated regimen: 6 fractions/week (total dose 66-70 Gy in 33-35 fractions) vs. conventional regimen: 5 fractions/week (same total dose)	LRC	5-year LRC was higher in the accelerated group compared to the conventional group. Higher acute morbidity was reported in the accelerated group, with no significant differences in late toxicity.
Lacas et al., 2021	Meta-analysis of randomized trials	To update results on the efficacy of concomitant chemotherapy in improving OS for non-metastatic HNSCC	Patients with non-metastatic HNSCC, included in trials from 1965 to 2016 (total 18,951 patients)	Main comparison: LRT vs. LRT + CT; Secondary comparison: Induction CT + radiotherapy vs. radiotherapy + concomitant (or alternating) CT	OS	Concomitant CT significantly improved OS. Efficacy decreased with age and poorer performance status. Induction or adjuvant CT did not improve OS.

Noronha et al., 2018	Phase III randomized trial	To assess the noninferiority of once-a-week cisplatin compared with once-every-3-weeks cisplatin in LA-HN SCC	LA-HN SCC (n = 300)	Once-a-week cisplatin (30 mg/m ²) vs. once-every-3-weeks cisplatin (100 mg/m ²), both given with concurrent RT	LRC	Once-every-3-weeks cisplatin resulted in superior LRC albeit with more toxicity, compared to once-a-week cisplatin. No significant difference in PFS or OS.
Rades et al., 2023	Retrospective cohort study with matched comparison	To investigate carboplatin as an alternative to cisplatin for CRT in patients with HNSCC who cannot receive cisplatin due to impaired renal function	Patients with HNSCC (n = 176; 131 receiving cisplatin, 45 receiving carboplatin)	Cisplatin and RT vs. carboplatin and RT	LRC	No significant differences in 2- and 3-year LRC, MFS, and OS. No significant differences in toxicities. Slightly more but insignificant number of patients completed carboplatin based chemotherapy compared to cisplatin group.
Gillison et al., 2019	Randomised trial	To determine if cetuximab can maintain high survival and reduce toxicity compared to cisplatin when combined with	Patients with HPV-positive OPCSCC (n = 849)	RT plus cetuximab vs. RT plus cisplatin	OS	Cetuximab did not meet non-inferiority criteria for OS. PFSI and LRC were significantly worse in the cetuximab group. Acute and late toxicities were similar between groups.

RT in HPV-positive OPCSCC						
Tao et al., 2020	Phase III randomized trial	To test the combination of avelumab with cetuximab and RT against SOC	LA-SCCHN (n = 82 in safety phase)	Cohort 1: SOC with IMRT and cisplatin (Arm A) vs. IMRT with cetuximab and avelumab (Arm B). Cohort 2: SOC with IMRT and weekly cetuximab (Arm D) vs. IMRT with cetuximab and avelumab (Arm C)	Safety phase monitoring grade ≥ 4 toxicity	The combination of avelumab, cetuximab, and RT was tolerable. Grade $\geq IV$ toxicity occurred in 12% of patients in experimental arms (all in Arm C), compared to 14% in Arm A and 10% in Arm D. Approval was given to continue the trial without modification.
Overgaard. 2011	Meta-analysis of 32 randomized clinical trials	To evaluate the impact of hypoxic modification techniques on outcomes in HNSCC treated with RT.	32 randomized clinical trials (totalling 4805 patients)	Various hypoxic modification techniques including normobaric oxygen, carbogen breathing, hyperbaric oxygen (HBO), hypoxic radiosensitizers, and combinations of such	LRC	Hypoxic modification of RT significantly improved LRC, DSS, and OS. No significant impact on DM or late RT toxicity. The benefit was consistent across different types of hypoxic modifications and even in trials using conventional fractionation schedules

Lassen et al., 2010	Retrospective analysis within the DAHANCA 5 trial	To assess the influence of p16-expression on the response to the hypoxic cell radiosensitizer nimorazole in HNSCC	Cohort from DAHANCA 5 trial (n = 331)	Nimorazole vs. placebo combined with RT	LRC	Nimorazole significantly improved LRC compared to placebo. HPV/p16-positive tumours may not benefit clinically from hypoxic modification due to reduced radioresistance.

Table 1.1: Summary of key evidence comparing treatment management options for patients with head and neck cancer. CRT = chemoradiotherapy; CT = chemotherapy; DAHANCA = Danish head and neck cancer group; DSS = disease specific survival; LA-HNC = locally advanced head and neck cancer; LRC = loco-regional control; LRFS = loco-regional free survival; LRT = loco-regional therapy; MFS = metastasis free survival; OPSCC = oropharyngeal squamous cell carcinoma; OS = overall survival; PFS = progression free survival; RT = radiation therapy; SCC = squamous cell carcinoma; SOC = standard of care.

1.2 RADIATION INDUCED XEROSTOMIA AND STICKY SALIVA

Despite advances in treatment planning and delivery, where the use of modulated techniques is now standard of care, up to 30% of patients with HNC still experience permanent xerostomia (37). As with all radiation induced reactions, xerostomia can be broadly divided into acute, sub-acute, and late toxicity. Acute toxicity occurs either during treatment or in the weeks immediately following treatment, and late toxicity describes events occurring months to years following treatment completion (38). Whilst no exact cut-off timepoint exists clearly separating these definitions, authors have described late xerostomia ranging anywhere from six weeks, to three months, or even 12 months following treatment (37, 39, 40). In a study assessing the relationship between salivary amylase and xerostomia, De Felice et al., (41) were even less specific defining acute and late xerostomia as events that simply occurred either during treatment or during follow-up. The authors were not clear when the “follow up” period commenced; however, it ranged from 15 to 40 months in duration. Although subjective sensations, acute and late xerostomia are linked with salivary flow rate and along with its associated conditions, such as sticky saliva and mucositis, has a recognised negative impact on patients’ quality of life (42, 43).

1.2.1 Physiology of the Major Salivary Glands and Their Response to Radiation

The major salivary glands include the parotid, sublingual and submandibular glands, and are responsible for the production of approximately 90% of the total saliva volume (44). The parotid gland, located in the retromandibular fossa is the largest of the three major glands, and under stimulated conditions secretes over 50% of the total saliva volume into the superior portion of the oral cavity via Stensen’s duct

(45-47). In contrast, under non-stimulated conditions, it is the submandibular glands which produces up to two-thirds of the saliva volume (46). The smallest of the three glands are the sublingual glands, and when stimulated, produce just 2-5% of the total salivary volume (48). The tissue structure of these glands also differs, as does the composition of the saliva they produce. Collectively the major salivary glands contain three main types of cells: acinar, ductal, and myoepithelial (Figure 1.1). The composition of the individual glands vary, specifically, the type of acini they contain; i.e. serous, mucinous, and seromucous (45). The sublingual glands predominantly contain mucinous cells, producing saliva that is viscous. The submandibular glands are of mixed composition containing both serous and mucinous components, yet the parotid glands are exclusively serous in nature (44). Whilst the saliva produced by the parotid glands is watery, with a neutral pH, it is rich in α -amylase; an enzyme critical for digestion of carbohydrates (44). In addition to the role this amylase-rich saliva plays in the first stage of digestion, the mucins secreted by the viscous saliva of the sublingual glands lubricate the oral cavity forming a protective barrier against infection, as well as supporting speech and taste sensation (46). The differing cellular composition of these three major glands was also thought to impact on their response to radiation, with Deasy et al., (47) highlighting the uncertainty regarding the radiosensitivity of the individual glands.

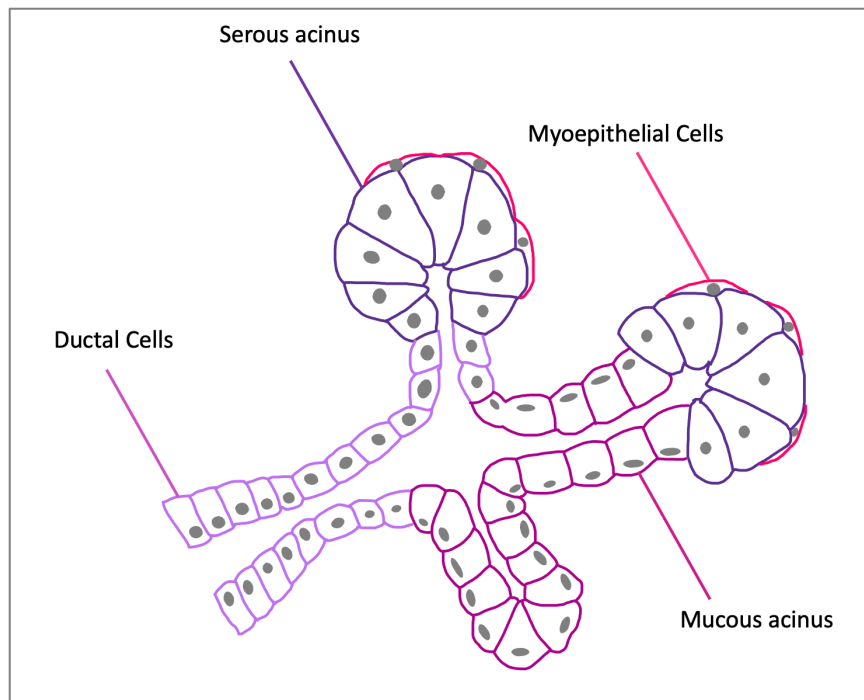


Figure 1.1: Schematic diagram of the cellular composition of salivary gland.

Based on the slow proliferation rate of the cells within the glands, theoretically they should be reasonably radioresistant (49). Despite this theory, salivary flow rate can be reduced by as much as 50%-60% within the first week of treatment (50). Similarly, Burlage et al., (51) reported functional loss as high as 80% in the first two weeks of treatment, indicating this tissue is acutely responsive to radiation. Several authors state that it is specifically the serous acinar cells within the glands that experience the greatest loss in response to radiation (49). Whilst this implies that the composition of the parotid glands may render them more radiosensitive compared to the submandibular and sublingual glands (52), this radiobiological theory also has not been observed in practice, indicating the complexity of the glands' microenvironment and the impact this cellular damage has on the progression of both acute and chronic xerostomia (51, 53). Consequently, the cellular organisation of the individual glands has been examined in greater detail, with the intention of identifying sub-regions of increased radio-sensitivity or radio-resistance (54, 55).

Irradiation of the salivary glands causes rapid acinar cell loss resulting in hypofunction of the gland and a decrease in salivary volume. This acute loss of acinar cells also influences chronic xerostomia downstream (i.e., 12 months following exposure to radiation). In this late phase, the impact of previous irreparable damage to the parotid parenchyma is complex and multifactorial (53). One theory is that persistent xerostomia is thought to be caused by a consequential reduction of the stem cells which occurred during treatment; the effect of which is an inability for the lost acinar cells to be replaced (54, 55). Limiting the dose to these progenitor cells could hypothetically support long term cell repair, promoting functional recovery of the gland (55). This theory has prompted the investigation into selective stem cell sparing radiotherapy, in which dose to the stem cell rich region surrounding the major ducts within the parotid is avoided. This approach incorporates results from animal studies which mapped the concentration of stem cells within the gland (55). Planners are then able to exploit the dosimetric flexibility afforded by modulated planning techniques to create intentionally heterogeneous dose distributions within the gland, whilst maintaining the same overall mean dose (47, 56). A double blind RCT published in 2022 demonstrated the feasibility of this method to treatment planning yet failed to meet the primary endpoint of the trial, i.e. no difference in prevalence of $\geq 75\%$ reduction in stimulated parotid gland saliva production between stem cell sparing and standard planning (56). Although the primary end point was quantified based on salivary flow, secondary endpoints included physician scoring and patient reported outcomes measures (PROMs) using several different rating scales measured at 6, 12, and 24 months following treatment, thereby focussing their attention on chronic xerostomia. This demonstrates the complexity of measuring radiation induced toxicities associated with parotid gland damage and highlights the variety of established approaches in clinical use.

1.2.2 Quantifying Radiation Induced Xerostomia and Sticky Saliva

Despite the cellular differences between the individual salivary glands and the radiobiological conundrum that surrounds them, the overall functional impact from the patients' perspective is similar. The decreased salivary flow rate is ultimately translated as xerostomia; a subjective sensation to describe the oral dryness experienced by patients. Not only is the quantity of saliva impacted, but so too is the composition of the saliva that is produced as a result of the decrease in serous secreting cells. The saliva pH decreases and becomes more viscous, expressed as "sticky saliva" from the patients' perspective (50). This stickiness is exacerbated by the inflammation of the salivary duct that occurs (53), as represented in the Common Terminology Criteria for Adverse Events (CTCAE) (57). This changing volume and viscosity put patients at an increased risk of dental caries as well as increased susceptibility to oral infections. Basic functions associated with daily living, such as mastication and swallowing are also impacted, as are taste and speech. Although the parotid gland produces most of the saliva under stimulated conditions, damage to the submandibular gland will impact on non-stimulated salivary production causing increased nocturnal oral discomfort (55). Assessing the impact of radiation on the parotid glands can be approached in different ways. This may include evaluating the radiation-induced damage from an anatomical or physiological perspective, or alternatively assessing the clinical symptoms that manifest as a result of this damage.

Sialometry is one method used to directly measure the salivary output within the whole oral cavity, or selectively measuring output from individual glands (58, 59). The latter approach relies on catheterisation of the specific duct (e.g. Stenson's duct in the case of the parotid gland) or using a Lashley cup placed at the duct orifice, as was done by Steenbakkers et al., (56). In this approach, saliva is collected over a set period of time based on either stimulated or non-stimulated conditions. Although this quantitative approach has been used in many previous radiotherapy studies (60, 61), results can be inconsistent due to lack of reproducibility (48, 58). Variability in results may be caused by patients eating, drinking,

or smoking before the procedure, and can even be influenced by circadian rhythms, as demonstrated by Ishikawa et al. (62, 63). In an attempt to account for the latter on salivary volume measured at 3, 6, and 12 months following radiotherapy, Murdoch-Kinch et al. (64), collected samples between 8am and 12am. Steenbakkers et al., (56) also attempted to mitigate some of these uncertainties by first drying the buccal mucosa with gauze, standardising equipment, and having the procedure performed by “well-trained” staff. Whether sialometry is performed based on whole salivary output or on one gland alone depends on the purpose of the test. If the objective is to quantify radiation dose response of the normal tissues, then targeted sialometry is required. Conversely, if the goal is to establish the association with clinical symptoms then whole output measurements are required due to the contribution of the minor salivary glands that are dispersed within the oral cavity are included (48). Performing saliva collection methods allows not just measurement of the total saliva volume produced, but also provides the opportunity to examine the salivary composition (48). Although not commonly performed, mucins and/or pH can be measured which may indicate risk of other potential complications, such as oral infection and dental caries. However, if using citrus based stimulants, the pH and composition of the saliva can be altered, and pH analysis under unstimulated conditions may be more appropriate (59). Despite the commonality and availability of sialometry, salivary flow rate is not always correlated with xerostomia, and alternative solutions must be considered (59).

Medical imaging can also be used to assess radiation induced changes to the salivary glands, and has the benefit of being able to assess xerostomia at any point during its progression. Depending on the imaging used, these changes detected can either be morphological or functional. Ultrasound (US) imaging may be attractive given the lack of ionising radiation; however, its use in the assessment of radiation induced salivary gland damage is limited. Ying et al., (65) used grey-scale and doppler ultrasound to describe the appearance of the parotid gland of patients who had received radiotherapy compared to a matched healthy population. Significant

differences were noted relating to the size of the gland as well as the brightness of the duct and vascular resistance. Despite these results, this study did not directly link these changes with the incidence or severity of xerostomia or sticky saliva. Operator dependency is a known limitation of ultrasonography, but in the case of assessing salivary glands other challenges must also be noted. The parotid glands of patients with Sjogren's syndrome show similar imaging characteristics on ultrasound making differential diagnosis with radiation induced tissue injury difficult. Furthermore, the inability to assess the deep lobe of the parotid gland due to interference of the mandibular ramus is particularly problematic (58).

Computer tomography on the other hand allows for complete visualisation of the glands, allowing for assessment of parotid gland shrinkage associated with acinar cell loss. Furthermore, CT is widely available and is an imaging type used routinely in radiotherapy. Despite this, there is a paucity of evidence to support its use for assessing salivary gland damage in the pre-radiomics era. This is presumably due to the lack of soft tissue contrast limiting detailed analysis of subtle glandular changes. Furthermore, the use of ionising radiation from CT imaging has also been highlighted as a limitation by some (58, 59).

With the lack of ionising radiation and the ability to visualise all the major salivary glands as well as their internal architecture, magnetic resonance imaging (MRI) is an attractive alternative to both US and CT. MRI will give excellent soft tissue contrast and the increased signal intensity of a T2 weighted sequence is indicative of salivary gland injury (66). Several studies have utilised MRI to highlight parotid volume reduction as well as signal intensity changes following radiotherapy to the head and neck (62). Whilst historically, MRI has been criticised due to its expense and inaccessibility, it is now included as part of routine follow up imaging protocols (67). Furthermore, the increased use of MR linear accelerators (MR-Linacs), and general availability of MRI scanners in radiotherapy departments provides an opportunity for longitudinal assessment of treatment response. In addition to standard sequences, researchers are

now also looking to the additional functional information provided by diffusion weighted MRI (DW-MRI). Reviewing the variation in apparent diffusion coefficient (ADC) based on DW-MRI acquired at baseline and 10 days into treatment, Duffton et al., (68) noted changing tissue characteristics of the salivary glands. Specifically, a decrease in volume and a significant increase in ADC value was observed for the parotid glands, indicating a change in the cellularity and vascularity in response to radiation. An additional observation from this study was the difference in ADC values between each of the major salivary glands at baseline. This finding confirms from a functional imaging perspective the differences between the glands' cellular composition.

Functional imaging of radiation induced salivary gland imaging is not confined to MRI, but also includes positron emission tomography (PET) and single photon emission tomography (SPECT). Nuclear medicine imaging (NMI) allows for physiological processes to be mapped using radioactive tracers and has been used for the assessment of treatment related toxicities in a number of organs at risk (OAR) (69). The review by Mohan et al., (69) provide an historical overview of the various radiotracers used to assess salivary gland toxicity. Specific tracers that have successfully correlated with either a dose-response relationship, sialometry results, or observer-rated toxicity outcomes include [^{99m}Tc]pertechnetate (used with either scintigraphy or SPECT), [^{18}F] Fluorodeoxyglucose (FDG), and [^{11}C]methionine. As an example, the metabolic clearance rate of [^{11}C]methionine was shown to be positively correlated with salivary output from the parotid glands (70). A follow up study by the same author then used [^{11}C]methionine to demonstrated the sigmoidal relationship of parotid gland function with radiation dose, suggesting a TD_{50} of 30 Gy. More recently, ^{68}Ga llium – prostate specific membrane antigen (PSMA) has been used as a marker of acinar cell loss on PET imaging. Mohan et al., (69) propose that the quantitative information revealed with these tracers could be used to establish new dose constraints and more personalised treatment plans using baseline functional imaging. Despite the progress in imaging techniques to assess salivary gland damage, sticky saliva and xerostomia are

clinical symptoms of functional loss. As such, questionnaires, either observer-based or PROMs, have been the mainstays of reporting these outcomes for many years.

Questionnaire based approaches to toxicity assessment have been around as early as the 1970s when the WHO established one of the first reporting systems for cancer therapy (71). During the decades that followed a number of different scoring systems were introduced, each with their own limitations. Many of the systems in circulation at the time also only considered single modality treatments, thereby undervaluing the combined effects of oncology therapies common in HNC management (72). One of the earliest radiotherapy specific toxicity scoring system is the RTOG)/European Organization for Research and Treatment of Cancer (EORTC), first published in 1995 (73). Although still commonly used, this system was criticised by Eisbruch et al., (48) due to the ambiguity regarding observer rated or patient rated events. With respect to xerostomia, the authors also felt there was a lack of clarity regarding if data collected are based on stimulated or unstimulated conditions.

Efforts in correcting some of these shortcomings has resulted in the CTCAE system used today, for monitoring both acute and late events (57). Whilst the current version of the CTCAE does not contain the term “xerostomia” explicitly, “dry mouth” is a CTCAE term characterised by reduced salivary flow in the oral cavity. The grading of this event is based on symptoms of dry mouth and thick saliva, and their impact on oral intake. Additionally, unstimulated salivary flow rates with pre-defined thresholds for grades 1-3 are detailed. Saliva thickness is reflected in the “dry mouth” item, and saliva stickiness is captured within the item pertaining to salivary duct inflammation. Sticky saliva is considered as grade 2 salivary duct inflammation. The CTCAE also considers impact on “activities of daily living” in an attempt to capture the patient’s quality of life; however, there is often a lack of concordance between the physician and the patient in this regard (74, 75). The risk of undervaluing the impact of treatment on quality of life has led to the development of separate questionnaires addressing

these concerns. One such example is the EORTC Quality of Life Questionnaire (EORTC QLQ-C30). This questionnaire consists of 30 items capturing QoL events relevant to all patients with cancer.

The widespread use of the EORTC QoL modules has shed light on patients' perspectives and has contributed to the expansion of PROMs used right throughout the survivorship period. One example is the PRO-CTCAE (76). This tool, used in conjunction with the standard CTCAE tool, captures the presence/absence, frequency, severity, and interference of events for items deemed appropriate for self-reporting. In addition, the electronic nature of the form allows for conditional branching logic to increase efficacy in reporting (77). There are also several site-specific PROMs and QoL questionnaires for HNC, such as Head and Neck module of the EORTC QLQ-C30 (QLQ H&N35) (78). These tools can be used in conjunction with others but allow more room for cancer specific considerations (79). Often using a combination of Likert scales and free text responses, several probing questions beyond simply "dry mouth" and "sticky saliva" are intended to comprehensively capture the breadth of difficulties associated with these conditions (80). For patients with HNC, questions pertaining to mastication, dysphagia, and talking are common examples that can help identify clinically meaningful toxicity.

Fully appreciating the multidimensionality of quality of life requires capturing the voice of the patients directly. Here, the use of PROMs are a welcome complement to observer-rated scales by facilitating open communication between the patients and health care professionals. The feedback provided by patients using these tools also allows for a proactive approach to managing patients' side effects (80).

1.2.3 Prevention and Management Xerostomia and Sticky Saliva

Preventative and management strategies of xerostomia and sticky saliva have developed over the years largely driven by technological and biological advances in radiation oncology. In the 2023 review by Nathan et

al., (81) these strategies can be loosely divided into pharmacologic and non-pharmacologic interventions.

Various pharmacological agents have been explored for the prevention of radiation induced xerostomia. Some examples include intraglandular injection of the botulinum toxin or the use of vascular endothelial growth factors; however, both are currently limited to animal studies (59). The use of antioxidants, such as amifostine, allows for scavenging of oxygen free radicals thereby acting as a tissue protectant to radiation without evidence of interfering with tumour response (59, 82, 83). Approved by the Food and Drug Authority (FDA) in 1999, the efficacy in reducing the incidence of moderate to severe acute xerostomia was highlighted in a Cochrane analysis published 18 years later (84). Despite this, concerns regarding drug related toxicity may have contributed to its limited widespread use. There have been reports in the literature of dermatitis, mucositis, hypotension, and vomiting (59); although, as stated in the meta-analysis by Gu et al., (83) if administered subcutaneously rather than intravenously these events may be minimised.

Non-pharmacologic interventions investigated take a less systemic approach, although in the case of submandibular transfer, can be invasive. The surgical relocation of the submandibular glands to either the submental space or even temporary re-housing in the patient's forearm has been successfully performed by several groups (81). Positive results were confirmed in a meta-analysis of 177 patients from seven studies in which, late xerostomia under both stimulated and unstimulated conditions was significantly reduced compared to the control (85). The authors rationalised the need for surgical approaches to salivary gland sparing given the limited availability and mixed results with modulated radiotherapy (85). This notion can be definitively refuted based on several more recent meta-analyses of RCTs (86, 87). When compared with 3D conformal radiotherapy (3DCRT), patients randomised to receive intensity modulated radiotherapy (IMRT) experienced significantly less grade 2 acute and late xerostomia. Results such as these is evidence of the clinical benefits

afforded by the dosimetric flexibility of IMRT. In comparison to 3DCRT, IMRT allows for greater control over the dose distribution in treatment planning with the use of an inverse optimisation algorithm. The resultant modulated fluence patterns of each beam are then delivered via dynamic multi-leaf collimators. The dosimetric gain from this approach is the ability to create isodose distributions which can conform to concave, complex target doses. This is particularly relevant for HNC in which multiple large target structures are treated simultaneously, surrounded by numerous radiosensitive organs at risk (OARs). One of the drawbacks of IMRT is the protracted beam on time and high number of monitor units (MUs), increasing treatment duration. Although intrafraction motion is less of a concern when treating HNC due to the stable internal anatomy and external immobilisation used, patient comfort must be considered. Furthermore, the cellular repair and redistribution that can occur during treatment delivery is another shortcoming of the extended beam on time associated with IMRT (88). These concerns have been somewhat alleviated with the use of volumetric modulated radiotherapy (VMAT), which allows for similar dosimetric and clinical gains as IMRT yet with a shorter delivery time and reduced MUs. Definitive chemoradiation using IMRT or VMAT is the current standard of care for patients with OPC, and despite these technological advances a significant proportion of patients still experience xerostomia and sticky saliva requiring supportive care (87).

As with preventative strategies, management of symptomatic xerostomia and sticky saliva can also include pharmacologic or non-pharmacologic approaches. Non-drug methods explored to date include acupuncture and electrical nerve stimulation via a Transcutaneous Electrical Nerve Stimulation (TENS) machine, neither of which have conclusive high-level evidence to support their use (81). Similar to the nerve stimulation from TENS, there are also some approved drugs that mimic saliva-inducing nerve signals, thereby increasing salivary flow (89). However, as mentioned by Sasportas et al., (89) whether this translates to a reduction in patient perceived xerostomia remains to be seen. Hypobaric oxygen therapy (HBOT) has also been investigated based on its mechanism to increase

oxygenation of the blood, supporting tissue repair (90). Using the EORTC QLQ H&N35 as an assessment tool, patients who received HBOT have demonstrated a negative correlation with sticky saliva and xerostomia scores (90). Despite these results, Li et al., (59) call for clarity over the number of treatments of HBOT required, as well as the optimal start time. Furthermore, HBOT may be cost prohibitive for many patients, prompting the need for alternative solutions. Synthetic salivary substitutes are cost effective solutions designed to mimic the viscosity, lubrication and antimicrobial properties of saliva (89). Although earlier studies indicated a reduction in intra-oral dryness in as many as 56% of patients (91), a randomised double-blind study published this year failed to demonstrate statistically significant improvements in xerostomia complaints or QoL (92). These results are consistent with a 2011 Cochrane review which failed to demonstrate efficacy of any topical agent compared to a placebo (93). These results demonstrate the inadequacy of salivary substitutes to replicate the complex composition of saliva described by Eisbruch et al., (48).

There is an abundance of evidence investigating the prevention and management of xerostomia and sticky saliva, as summarised in Table 1.2. Anticipating these events, however, will better prepare patients and facilitate prompt action by health care providers. To date the prediction of toxicity risk relies on evidence derived from seminal modelling studies.

	Approach	Intervention	Comments
Prevention of Xerostomia	Pharmacologic	Intraglandular injection of Botulinum toxin	Evidence limited to animal studies
		Vascular Endothelial Growth Factors	Evidence limited to animal studies
		Antioxidants	Amifostine was FDA-approved in 1999, scavenges oxygen free radicals, efficacy confirmed in Cochrane analysis, concerns regarding toxicity.
	Non-pharmacologic	Submandibular Gland Transfer	Surgical relocation to submental space or forearm (positive results in meta-analysis).
		Radiotherapy Techniques	IMRT: Less grade 2 acute and late xerostomia compared to 3DCRT. VMAT: Similar benefits to IMRT, shorter delivery time.

Management of Xerostomia	Pharmacologic	Drugs Mimicking Nerve Signals	Increases salivary flow, effectiveness on patient-perceived xerostomia uncertain.
		Hypobaric Oxygen Therapy (HBOT)	Increases blood oxygenation, mixed results, cost concerns.
		Synthetic Salivary Substitutes	Cost-effective, limited efficacy.
	Non-pharmacologic	Acupuncture	Explored for symptomatic relief, lacking conclusive high-level evidence.
		Electrical Nerve Stimulation (TENS)	Similar to nerve stimulation, lacking conclusive high-level evidence.

Table 1.2: Summary of the pharmacological and non-pharmacological interventions that have been explored in the prevention and management of parotid toxicities.

1.3 MODELLING NORMAL TISSUE DAMAGE

Normal tissue complication probability (NTCP) models attempt to reduce a multitude of dosimetric and clinical variables into a single value which will estimate the risk of toxicity for a given organ (94). More recently NTCP models have also been used for patient selection for proton therapy in which patients qualifying for proton therapy was based on a risk assessment of the likely xerostomia and dysphagia incidence (95). Before applying NTCP models into the clinic, we must first be aware of inherent limitations associated with them. At the time of publication of the quantitative analysis of normal tissue effects in the clinic (QUANTEC) papers, Marks et al., (94) highlighted several caveats to the models presented within the individual papers. Primarily, synthesizing literature and pooling data from the evidence base was challenging, largely attributed to the incomplete results presented or ambiguous endpoints used (96). Toxicity endpoints used in the pre-QUANTEC era were dominated by observer-rated scores, yet with the increasing use of PROMs in recent years there is now a need to re-visit recommendations (97). In addition, treatment related factors also impact on their clinical utility, for example the radiotherapy fraction size as well as the use of combined treatment modalities. This point is relevant today given the evolution in management of head and neck cancer which now utilises multimodality treatments. Finally, the evidence underpinning the QUANTEC recommendations was derived based on 3DCRT. Today IMRT and VMAT are standard of care for head and neck radiotherapy and the dose profile these plans produce is vastly different. Dose across the organs at risk is heterogeneous and the simplicity of the DVH based NTCP models lack spatial information regarding this complex distribution.

Irrespective of these shortcomings, the NTCP models reported in the QUANTEC papers, and the dose volume constraints (DVCs) presented have been widely adopted in radiotherapy around the world. Historically these

models were generated based on logistic regression, and Marks et al., (94) categorises them under three sub-headings; i.e. DVH-reduction models, tissue architecture models, and multiple-metric models. The DVH-reduction models estimate the dose which will induce a complication in 50% of the population (TD50). These estimates are based off an assumption of an homogenous dose distribution across the volume; however, their application in clinical practice is often to structures which have a non-uniform dose distribution (94). This limitation was noted at the time of publication in 2010 and is even more relevant in the current era of modulated techniques which create rapid dose fall off across structures (98). To overcome this, the portion of the volume irradiated was integrated into the model using mathematical assumptions, such as the Lyman assumption and the generalised uniform equivalent dose (gEUD). When used together, this is referred to as the Lyman-Kutcher-Burman (LKB) model. This model has since been used in several studies to estimate the risk of xerostomia and has also been integrated into commercial treatment planning systems facilitating biological based dose optimisation and plan evaluation (99-101). Despite their widespread use in clinical practice, Issakson et al., (102) casts doubt over these mechanistic models and their ability to predict radiation induced toxicity. This criticism is plausible considering the persistent incidence of xerostomia, even when modulated techniques are used and dose received is deemed to be acceptably low (43). Issakson et al., (102) criticise the complexity of the “handcrafted” algorithms used within the current NTCP models, as well as the oversimplification of the interaction between radiation dose and normal tissue damage. Irrespective of the root cause, the imperfection of these models has also been recognised by others (103, 104). Despite these caveats, when fitting their clinical data to the LKB model, Dijkema et al., (101) were able to confirm the parallel architecture of the parotid glands. This concept of tissue “architecture” explains the different radiosensitivity of individual organs at risk based on the arrangement of their functional sub-units and is referred to as “tissue architecture” models by Marks et al., (94, 105). This implies that small volumes of the parotid gland can

withstand receiving high doses. The multi-metric model described by Marks et al., (94) suggests that relying on a single point along the DVH curve may be overly simplistic and that assessing several points along the curve may be more clinically meaningful. A single DVC may be appropriate for organs in which their functional subunits are arranged in serial architecture; however, for larger parallel organs the whole DVH curve may be more meaningful. Investigating the clinical viability of this is facilitated by machine learning approaches which can handle large amounts of data (102). Hansen et al., (98) investigated this approach to build an NTCP model to predict mucositis in a cohort of 802 patients with head and neck cancer. Using principal component analysis (PCA), they reduced the complete data extracted from DVH curve of the oral cavity down to five principal components. Feeding these into their model along with several clinical predictors, they concluded that treatment acceleration and mean dose to the oral cavity were the two most dominant predictors for oral mucositis. This study is formidable in that it assesses the entirety of the DVH curve of the extended oral cavity. Unfortunately, the disregard for the dose to the salivary glands as separate dosimetric predictors is disappointing given the established relationship between salivary flow and mucositis (43).

The imperfection of many of the existing NTCP models and the lack of adequate predictors was “unsettling” according to Marks et al., (94) prompting continued efforts in this field. Although the more recent work of Hansen and others have enhanced toxicity prediction by exploring additional covariates, they are often still confined to clinical and dosimetric variables (98, 106). With this approach, data contained within medical images continues to remain unseen and undervalued. Quantitative analysis of medical images with radiomics could reveal underlying biology, essentially uncovering clinically meaningful imaging biomarkers to enhance toxicity prediction for patients.

“Radiomics” is the process of extracting large amounts of quantitative data from images that can be used for data mining and subsequent analysis (107). Combining this data with clinical information can potentially predict treatment outcomes and assist in clinical decision making. The application of radiomics to oncology is based on the premise that specific radiomic features are reflective of the tissue’s heterogeneity within a region of interest, essentially revealing detail regarding the underlying tissue biology (108). When referring to the tumour volume in this regard, some authors liken radiomics to a non-invasive “digital biopsy” (109). A recent review by Tortora et al., (110) highlighted the breadth of studies which used radiomics to examine the tumour volume within the head and neck region. As one example, the potential of radiomics to classify tumours based on HPV expression has been demonstrated by Bagher-Ebadian et al., (111) among others, thereby overcoming some of the challenges with traditional immunohistochemistry (112-114). In addition, several other studies included in the review by Tortora et al., (110) successfully used radiomics to grade and stage primary tumour volumes and positive regional lymph nodes. Radiomics also has the potential to reveal detail regarding the radiosensitivity of organs at risk. This has been applied to multiple structures within the head and neck region, such as the mandible to predict osteoradionecrosis (115), or the parotid glands to predict xerostomia (116).

Whether applied to a tumour volume or organ at risk the detail revealed with radiomics will yield important information regarding the underlying tissues and their response profile to radiation (108, 117). Radiomics has been validated for a range of imaging modalities many of which are used routinely in head and neck radiation oncology; including, MRI (118), PET (119), CT (120) and CBCT (121). When combined with curated clinical data, radiomics can act as a decision support tool to facilitate personalisation of treatment and patient management.

1.4.1 Radiomic Features

There are several commercially available software solutions to support extraction of radiomic features, such as Imaging Biomarker Explorer (IBEX), Local Image Feature Extraction (LIFEx) or PyRadiomics (122-124). Alternatively, researchers can develop codes in-house to perform the task. Given the variability in the systems that exist, and the influence on individual values extracted, the imaging biomarker standardisation initiative (IBSI) was established (125). As the name suggests, this international collaboration serves to standardise the feature extraction process including the data processing workflow. Furthermore, they serve to provide a clear set of relevant definitions and nomenclature, thereby facilitating transparent dialogue amongst researchers. As a result of their efforts, we can describe the fundamental types of features extracted that are common between systems (126). In addition to the IBSI, several reviews and commentary papers pertaining to radiomics have summarised the features extracted (108, 127, 128).

In 2016, Gilles et al., (127) broadly described the two main types of radiomic features as semantic and agnostic features. Semantic features include those which have been used by radiologists to describe imaging results, such as the volume, sphericity, and compacity of the region of interest (ROI). The role of radiomics here is to simply extract this information in a more objective, computer assisted fashion. Agnostic features on the other hand describe features that are mathematically extracted from the image and extend beyond standard imaging descriptors (127). Agnostic features can be further divided into three main statistical sub-groups; first order, second order and higher order (128), as depicted in Figure 1.2.

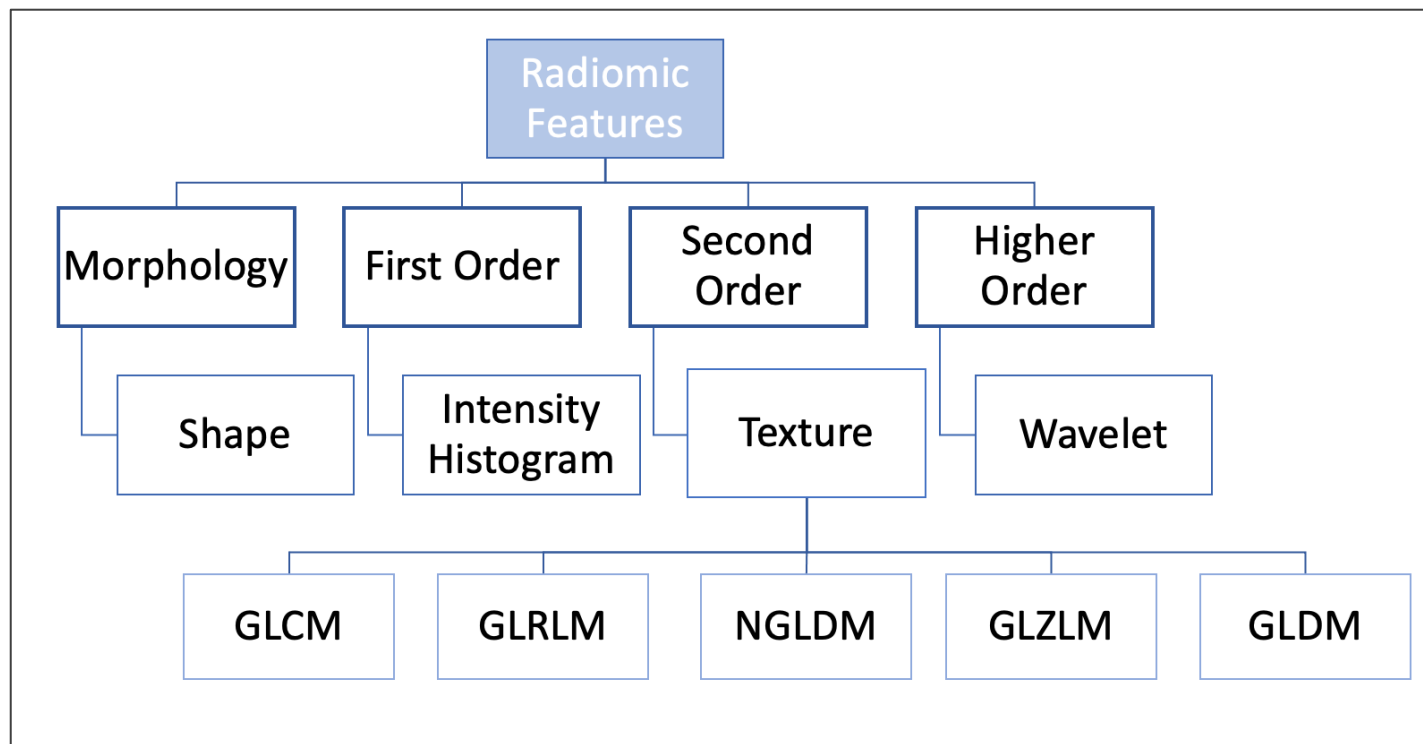


Figure 1.2: Classes of radiomic features. Texture feature families include Grey level co-occurrence matrix (GLCM), Grey-level run length matrix (GLRLM), Neighbourhood grey-level different matrix (NGLDM), Grey-level zone length matrix (GLZLM), and Grey Level Dependence Matrix (GLDM)

First order statistics include descriptions of the intensity values of imaging voxels without consideration of spatial distribution within the ROI, and are typically referred to as histogram features (128). In CT imaging these features are calculated based on the global grey-level values derived from the CT numbers or Hounsfield units (HU) within the ROI (108). These data are then reduced to simple numerical values including the mean, median, minimum, maximum values, and percentiles within the ROI. In addition, the shape of the intensity histogram is described using symmetry (skewness) and flatness (kurtosis) of the histogram.

Second order statistics derived from radiomics are referred to as texture features. These features describe the spatial relationship of voxel intensities relative to their neighbour. This is done by assigning a numerical value to the relationship between similar or dissimilar contrast values of voxels across a volume. Describing the heterogeneity within the ROI, these features have been shown to be highly correlated with tissue heterogeneity at a cellular level (108). Within the hypernym of texture features, there are several feature “families” each containing various individual radiomic features. The Grey Level Co-occurrence Matrix (GLCM) considers the spatial relationship of pairs of voxels. In the LIFEx software, six features are calculated describing various paired relationships, for example contrast which represents the grey level difference between voxels within the pair (108, 123). The Grey Level Run Length Matrix (GLRLM) family describes the spatial distribution of runs of voxels which have the same grey value. Nardone et al., (129) puts this simply as the size of homogenous runs for each grey level within a ROI. In the Grey Level Size Zone Matrix (GLSZM) family, voxels are grouped if they have the same grey level as their neighbouring voxel (130). The number of zones within the ROI is then counted, and features within this family describe the size of these homogenous zones. A wider, flatter matrix indicates a more homogenous texture pattern of the ROI (108). Grey Level Distance Zone Matrix (GLDZM) is another variation of the GLSZM family; however, is not included in the LIFEx package. Mayerhoefer et al., (108) state that features within the GLDZM are essentially a hybrid of shape-based features and texture

features. This implies there may be collinearity of GLDZM features and others extracted with LIFEx. For this reason, the impact of this omission by LIFEx is likely to be inconsequential in modelling studies where feature reduction is required to minimise collinearity, although this theory remains to be tested. The Neighbourhood Grey Tone *Difference* Matrix (NGTDM) family has also been described by the IBSI (130). Within LIFEx, this family is referred to as the Neighbourhood Grey Level Difference Matrix (NGLDM) and reflects the differences between the grey level of one voxel and its neighbours within a set distance (108). With LIFEx software, this predefined distance includes 26 neighbour voxels in 3 dimensions. The final texture feature family detailed by the IBSI is the Neighbourhood Grey Level *Dependence* Matrix. Features represented here capture the overall coarseness of a ROI. Similar to the IBSI NGTDM family, features calculated in the Neighbourhood Grey Level *Dependence* Matrix consider their neighbourhood but are now considered connected to the central pixel (108, 130). This family, also not captured within LIFEx software is essentially an alternative to the GLCM family (130).

Finally higher order statistic features refer to the features derived from placing a filter over an image to repetitive or varied patterns within the region of interest (131). These features, commonly referred to as transform-based features, are also not represented in the LIFEx software.

Not only are there differences in the feature families included in different software options, but the mathematical method underpinning how individual features are calculated can also vary. A prime example of this is the difference in the Contrast feature of the NGLDM family. The formula used to calculate this feature in LIFEx software is based on Amadasun and King (132) in which the number of voxels within the neighbourhood is used as the denominator in the equation; however, if calculated using PyRadmics, this value is squared instead (133). These differences in calculation methods impacts the final feature value but also impacts on direct comparison between studies which use different software. Whilst software developers simply caution care should be taken in this regard,

Bibault et al., (126) raise concern regarding reproducibility of radiomic studies. Reproducibility of radiomics is influenced not just by the actual feature calculation methods, but also several variables introduced within the radiomics pipeline. These challenges have been discussed in detail by several authors, such as Gilles et al., (127) and Rizzo et al., (128).

1.4.2 Radiomics workflow

Whilst the clinical value of radiomics research continues to grow, an understanding of the process and potential pitfalls is vital when conducting such research responsibly (128).

The radiomics pipeline follows a distinct workflow which includes the following steps (134, 135):

1. Image acquisition and reconstruction
2. Image pre-processing
3. Region of interest (ROI) segmentation
4. Feature extraction
5. Feature reduction/selection
6. Data analysis, including correlation with clinical information and prediction of outcomes

Each component of this process presents its own challenges, and features may be influenced by variations in each step (127, 135). Investigation into these individual elements is an important step prior to embarking on radiomics research, in which clinical conclusions may be drawn based on unstable or inappropriately extracted features.

Image acquisition and reconstruction parameters vary considerably based on imaging protocols used. These inter-institutional disparities can introduce noise in the patient cohort, masking underlying biology represented by the radiomic feature values (117). In a systematic review by Reiazi et al., (136) they labelled this phenomenon as “image perturbation factors”. They went on to group relevant imaging characteristics from the

literature into three categories: scanner parameters, image acquisition parameters, and image reconstruction parameters. In their concluding remarks, they recommended standardising image acquisition and reconstruction parameters; although they conceded this may not be a practical solution. Alternatively, they suggest correcting for the scanner parameters during image analysis. From a workflow perspective this can be achieved by standardising some imaging parameters prior to feature extraction, such as the voxel size, or alternatively, harmonising extracted features via mathematical post-processing (108).

Once images are acquired and prior to feature extraction, the ROI must be delineated. The challenges which clinicians and researchers have faced with delineation also impact on radiomics research. Manual segmentation is extremely time consuming and is particularly relevant when working with the large datasets required for robust radiomics work. In addition, the inter and intra-observer delineation variability will introduce bias downstream as individual features may be influenced differently by this. The impact this variability causes on radiomic feature values is not well understood and, as identified by Traverso et al., (137) details provided regarding ROI segmentation is often insufficient in radiomic studies.

Radiomic feature extraction itself is also not immune to variations in methodology. As previously discussed, the specific software used will influence the number and types of features extracted. Additionally, this will also impact the specific feature values themselves based on the mathematical algorithms within the software. There are limited studies that compare radiomics software head-to-head; however, inter-software differences are recognised (126). The difference in algorithms used for feature extraction will impact on the reproducibility of features across platforms, and as such transparent reporting of specific radiomics software is called for, albeit rarely done (137).

The final challenges that are faced by radiomics researchers are inherent to all modelling studies. Radiomics converts medical images to highly-dimensional data (127). Managing collinearity through feature reduction

and feature selection is necessary prior to generating a predictive model. Dimensionality reduction techniques can be performed using supervised or unsupervised methods, and whilst not flawed, can be computationally complex. The requirement of this process is to avoid producing a model that suffers from overfitting. Once the predictive model is generated, testing generalizability will support its application to a clinical workflow. The process of external validation is made possible when working with large datasets; however, accessing sufficiently large datasets can be challenging. This is particularly valid if data available is limited to single institutions, or when medical records are incomplete. Accessing and curating a sufficiently large data set including images and high-quality clinical parameters is a challenge when working with real-world data (117).

Radiomics, although computationally complex, provides an opportunity to supplement existing NTCP models by exploring additional predictors. Several authors have cautioned against relying on traditional NTCP models to predict radiation induced toxicities (102-104). Not only are these models overly simplistic in the limited number of predictors considered, they also fail to account for biological differences between patients. Quantitative imaging analysis allows us to reveal patient specific biology related to the ROI, represented by radiomic features. By integrating these radiomic features into toxicity modelling, we are exploiting the personalised data contained within routine medical images. Whilst the evidence supporting the clinical application of radiomics is ever-growing, we must remain vigilant to methodological flaws which could threaten progress. This thesis serves to contribute to the evidence based regarding radiomic-driven NTCP models in order to better predict incidence of xerostomia and sticky saliva following radiotherapy to the head and neck.

1.5 CENTRAL HYPOTHESIS AND AIMS OF THIS THESIS

This thesis addresses the limited availability of tools available for a more personalised approach in the assessment of parotid damage and prediction of xerostomia and sticky saliva in patients treated with chemoradiation for head and neck cancer. The central hypothesis is that the quantification of morphological and textural changes within the parotid gland using radiomics features extracted from CT imaging can predict acute radiation induced xerostomia and sticky saliva. This hypothesis will be tested across four specific aims:

1. To quantify the variability in the values of radiomic features extracted from a right parotid region of interest delineated by a series of independent observers.
2. To quantify the impact of spatial resampling on radiomic features reproducibility associated with CT scanners and scan protocols heterogeneity within a multi-institutional cohort of patients with head and neck cancer
3. To improve the prediction of acute xerostomia and sticky saliva through the inclusion of radiomic analysis of the parotid glands for patients with head and neck cancer.
4. To quantify the shape, intensity, and textural changes of the parotid gland that occur during a course of radiotherapy to the head and neck region.

2.1 PATIENT COHORTS AND RESEARCH ETHICS

This thesis consisted of four experimental studies, each of which used a different cohort of patients diagnosed with head and neck cancer. Studies 2 and 4 utilised data available on the open access Cancer Imaging Archive (TCIA), and as such did not require ethical approval (138-140). Established in 2011 and funded by the National Cancer Institute (NCI) Cancer Imaging Programme (CIP), TCIA is a repository of radiology and histopathology imaging. The publicly available archive of these medical data further facilitates research by also providing relevant supporting data such as patient demographics, treatment details and oncological outcomes as supplementary material (138).

Ethical approval was sought from the Trinity College Dublin (TCD) School of Medicine Research Ethics Committee (SOMREC) for Studies 1 and 3, which required the transfer of irrevocably anonymised computer tomography (CT) images, radiotherapy (RT) Structure Sets (RTSTRUCT), and specific patient information. For Study 1 data were provided by the European Society of Radiotherapy and Oncology (ESTRO), in which radiotherapy professionals attended an online Fellowship in Anatomic delineation and CONtouring (FALCON) workshop (141). These live, hands-on delineation workshops serve to provide interactive teaching of target and organ at risk contouring for radiotherapy planning. The data provided for this study included the CT image of a patient with oropharyngeal cancer used in this workshop along with the RTSTRUCT files of the expert and the participants who attended this workshop. Access to these data extracted from the ESTRO FALCON platform were shared by a nominated gatekeeper in the ESTRO office to TCD.

Supported by the Ulysses Award from the Irish Research Council (Award no. 211863.17055), a collaboration was established between researchers at Gustave Roussy, Villejuif, France and TCD. These funds facilitated site visits for the research teams and supported the transfer of data used for Study 3. The principle of data minimisation was respected, in which data were only collected and made available for analysis if they were deemed relevant to the study aims and objectives. With the support of a nominated gatekeeper at Gustave Roussy, these data shared with TCD included patient demographics, chemoradiation treatment details, dosimetric results, toxicity outcomes, CT images acquired at baseline and RTSTRUCT sets generated for treatment planning purposes. To identify and minimise potential data protection risks when conducting health research, a data protection impact assessment was conducted for this study and in addition to TCD SOMREC approval, local ethical approval from Gustave Roussy was handled by the French research partners.

The sample size of each study varied depending on the data available. Study 1 included one patient with 40 observers delineating the region of interest. Study 2 included two cohorts with 127 and 17 patients in each one. Study 3 included a cohort of 98 patients treated at Gustave Roussy under the REACH clinical trial (30), and Study 4 consisted of 26 patients. Of note, the 17 patients in the second cohort in Study 2 was a sub-set of the full patient cohort used in Study 4. Whilst every patient in this thesis received radiotherapy for head and neck cancer, the diagnosis ranged in the tumour stage and site within the head and neck region, as detailed in Chapters 4, 5, and 6.

The radiomics workflow as applied to all studies in this thesis is illustrated in Figure 1. To obtain information regarding the CT image acquisition and reconstruction parameters, metadata were extracted from the image DICOM header. The regions of interest, i.e., the whole left or right parotid glands were contoured by workshop participants for Study 1, and atlas based auto-contouring methods (Varian SmartSegmentation, v15.5) for Study 2. For Study 3 and 4, the RTSTRUCT sets were provided from the clinical departments, and the whole left and right parotid gland contours were left unchanged from what was generated by the treating radiation oncologist for treatment planning purposes. For each study contouring of the whole parotid glands followed the consensus guidelines published by Brouwer et al., (142).

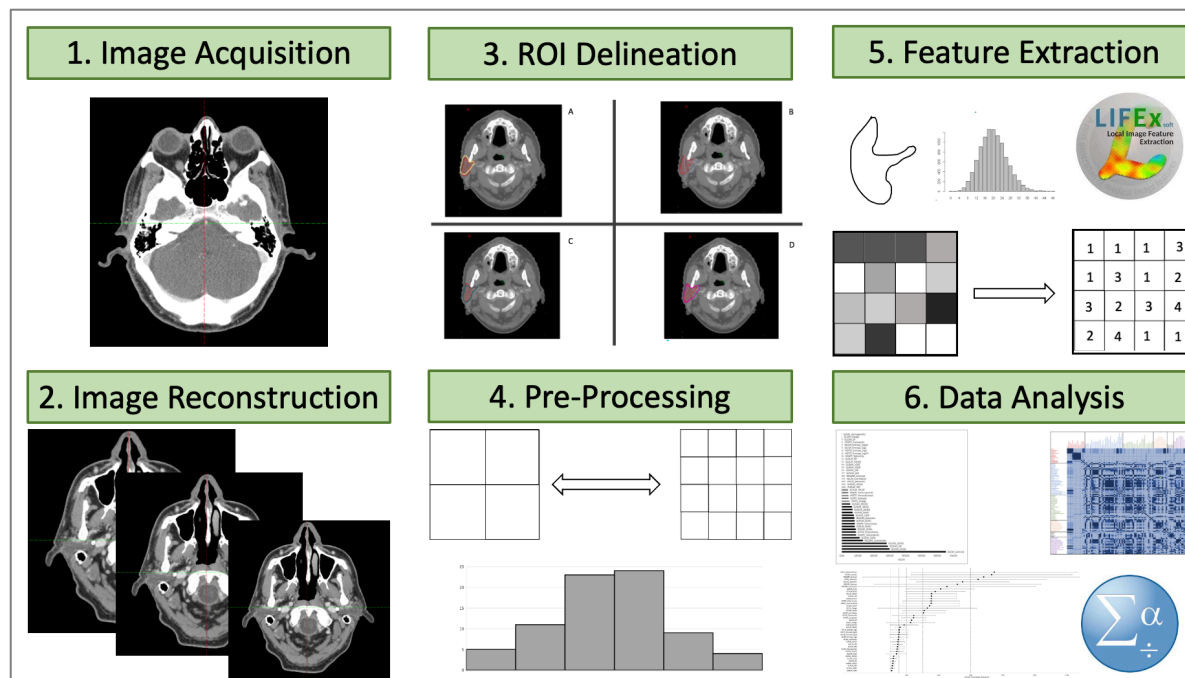


Figure 2.1: Radiomics workflow including (1) image acquisition; (2) image reconstruction; (3) region of interest (ROI) delineation; (4) pre-processing, including image spatial resampling and grey level discretisation; (5) radiomic feature extraction; (6) data analysis such as prediction of clinical outcomes using predictive modelling.

All image pre-processing and feature extraction were conducted using the Imaging Biomarker Standardisation Initiative (IBSI) compliant, open access Local Image Feature Extraction (LIFEx) software (123, 125). First released in 2016, LIFEx enables the extraction of conventional shape and histogram intensity information from medical images, as well as higher order texture-based analysis for the region of interest. The software supports the use of CT, ultrasound, magnetic resonance imaging (MRI), positron emission tomography (PET), and single photon emission computer tomography (SPECT). In addition to its ability to work with such a range of imaging types, the software is compatible with Windows, Linux and MacOS. The software also requires no programming skills from the user, thereby making it extremely accessible for medical imaging professionals without a background in computer science.

LIFEx has gone through several upgrades since its creation and the most up to date version of the software was used at the time each study was conducted. These software upgrades have impacted on the number of radiomic features extracted which vary based on the version used (Tables 2.1 and 2.2). The biggest change seen in this regard was the inclusion of IBSI labels in version 7.3, as well as the significant increase in intensity histogram features now available.

Feature Family	Feature Name	IBSI Code	Ch 1 v5.1	Ch 2 v6.2	Ch 3 v7.3	Ch 4 v6.2
Shape / Morphological	Volume	RNU0				
	Approximate Volume	YEKZ				
	Voxels Counting	-				
	Surface Area	C0JK				
	Surface To Volume Ratio	2PR5				
	Compacity	-				
	Compactness1	SKGS				
	Compactness2	BQWJ				
	Spherical Disproportion	KRCK				
	Sphericity	QCFX				
	Asphericity	25C7				
	Centre of Mass Shift	KLMA				
	Maximum 3D Diameter	L0JK				
	Integrated Intensity	99N0				
Intensity Histogram	Mean(HU)	X6K6				
	Variance(HU)	CH89				
	Skewness(HU)	88K1				
	Kurtosis(HU)	C3I7				
	Excess Kurtosis	-				
	Median(HU)	WIFQ				
	Minimum Grey Level(HU)	1PR8				
	10thPercentile(HU)	GPMT				
	25thPercentile(HU)	-				
	50thPercentile(HU)	-				
	75thPercentile(HU)	-				

90thPercentile(HU)	OZ0C				
Standard Deviation(HU)	-				
Maximum Grey Level(HU)	3NCY				
Mode(HU)	AMMC				
Interquartile. Range(HU)	WR0O				
Range(HU)	5Z3W				
Mean Absolute Deviation(HU)	D2ZX				
Robust Mean Absolute Deviation(HU)	WRZB				
Median Absolute Deviation(HU)	4RNL				
Coefficient of Variation(HU)	CWYJ				
Quartile Coefficient of Dispersion(HU)	SLWD				
EntropyLog10(HU)	-				
EntropyLog2(HU)	TLU2				
Area Under Curve Csh(HU)	-				
Uniformity(HU) (=Energy (V7))	BJ5W				
Root Mean Square(HU)	-				
Maximum Histogram Gradient(HU)	12CE				

Maximum Histogram Gradient Grey Level(HU)	8E60	   
Minimum Histogram Gradient(HU)	VQB3	   
Minimum Histogram Gradient Grey Level(HU)	RHQZ	   

Table 2.1: Shape/Morphological and Intensity Histogram radiomic features extracted per study based on specific version of LIFEx used. Green spot indicates feature was extracted; red spot indicated feature was not extracted. IBSI = Imaging Biomarker Standardisation Initiative

Feature Family	Feature Name	IBSI Code	Ch 1 v5.1	Ch 2 v6.2	Ch 3 v7.3	Ch 4 v6.2
GLCM	Joint Maximum	GYBY	●	●	●	●
	Joint Average	60VM	●	●	●	●
	Joint Variance	UR99	●	●	●	●
	Joint Entropy Log2 (=Entropy_log2 (V7.3))	TU9B	●	●	●	●
	Joint Entropy Log10 (=Entropy_log10 (V7.3))	-	●	●	●	●
	Difference Average	TF7R	●	●	●	●
	Difference Variance	D3YU	●	●	●	●
	Difference Entropy	NTRS	●	●	●	●
	Sum Average	ZGXS	●	●	●	●
	Sum Variance	OEEB	●	●	●	●
	Sum Entropy	P6QZ	●	●	●	●
	Angular Second Moment (=Energy (V7.3))	8ZQL	●	●	●	●
	Contrast	ACUI	●	●	●	●
	Dissimilarity	8S9J	●	●	●	●
	Inverse Difference (=Homogeneity (v7.3))	IB1Z	●	●	●	●

	Normalised Inverse Difference Inverse Difference Moment	NDRX				
	Normalised Inverse Difference Moment	WF0Z				
	Inverse Variance	WF0Z				
	Correlation	E8JP				
	Autocorrelation	NI2N				
	Cluster Tendency	QWB0				
	Cluster Shade	DG8W				
	Cluster Prominence	7NFM				
		AE86				
GLRLM	Short Runs Emphasis	22OV				
	Long Runs Emphasis	W4KF				
	Low Grey Level Run Emphasis	V3SW				
	High Grey Level Run Emphasis	G3QZ				
	Short Run Low Grey Level Emphasis	HTZT				
	Short Run High Grey Level Emphasis	GD3A				

	Long Run Low Grey Level Emphasis	IVPO	●	●	●	●
	Long Run High Grey Level Emphasis	3KUM	●	●	●	●
	Grey Level Non Uniformity	R5YN	●	●	●	●
	Run Length Non Uniformity	W92Y	●	●	●	●
	Run Percentage	9ZK5	●	●	●	●
NGTDM	Coarseness	QCDE	●	●	●	●
	Contrast	65HE	●	●	●	●
	Busyness	NQ30	●	●	●	●
	Complexity	HDEZ	●	●	●	●
	Strength	1X9X	●	●	●	●
GLSZM	Small Zone Emphasis	5QRC	●	●	●	●
	Large Zone Emphasis	48P8	●	●	●	●
	Low Gray Level Zone Emphasis	XMSY	●	●	●	●
	High Gray Level Zone Emphasis	5GN9	●	●	●	●
	Small Zone Low Grey Level Emphasis	5RAI	●	●	●	●
	Small Zone High Grey Level Emphasis	HW1V	●	●	●	●

Large Zone Low						
Grey Level	YH51					
Emphasis						
Large Zone High						
Grey Level	J17V					
Emphasis						
Grey Level Non	JNSA					
Uniformity						
Normalised Grey						
Level Non	Y1RO					
Uniformity						
Zone Size Non	4JP3					
Uniformity						
Normalised Zone						
Size Non	VB3A					
Uniformity						
Zone Percentage	P30P					
Grey Level	BYLV					
Variance						
Zone Size	3NSA					
Variance						
Zone Size Entropy	GU8N					

Table 2.2: Texture radiomic features extracted per study based on specific version of LIFEx used. Green spot indicates feature was extracted; red spot indicated feature was not extracted. IBSI = Imaging Biomarker Standardisation Initiative

Despite the differences between software versions, the fundamental workflow within the LIFEx interface remains largely unchanged. As van Timmeren et al., (135) highlight, several factors within the radiomics

workflow influence the feature values extracted. Within LIFEx, prior to feature extraction the user can specify pre-processing imaging parameters; including spatial resolution, image discretisation, and image intensity range.

Prior to feature extraction, it is recommended to resample the images to a common voxel size (108). This process essentially involves interpolating the images such that they have isotropic voxel spacing, which serves to reduce variation between datasets as well as ensuring the texture features are rotationally symmetric, or invariant (135). Within LIFEx this is achieved through polynomial lagrangian interpolation and will result in voxels being either up-sampled or down-sampled depending on their native resolution (123). This process was not conducted for Study 1 as there was only one patient with one CT scan included in the study. One of the aims of Study 2 was to investigate the impact of spatial resampling and as such radiomic features were extracted on both the native images as well as those resampled to a common voxel size of 1x1x1mm. For Studies 3 and 4, prior to radiomic feature analysis, all images were resampled to a common voxel size of 1x1x1mm.

Image discretisation is the process of arranging continuous quantitative data within the CT images into discrete bins. This reduces the calculation complexity ensuring feature calculation is tractable; i.e. the feature calculation is more controlled and as a result more manageable (135). As part of the discretisation process, either the width of the bins or the number of bins can be set, with the former approach being taken for each study in this thesis. In Study 1, the default setting of 10 HU in LIFEx was employed; however, testing the reproducibility of radiomic features was one of the aims of Study 2. Radiomic features were, therefore, extracted using both a bin size of 10 HU and 5 HU. Both bin widths were also then used for feature extraction in Study 3 to assess the reproducibility of study results when clinical outcomes were considered. Finally, based on observations from Studies 2 and 3, for Study 4, a fixed bin width of 5 HU was used.

A final step of the pre-processing stage is to set the intensity range in which the number of bins will be determined. For radiomics based on CT images, this involves limiting the range of HU “seen” by the algorithm to pre-set values, and is useful for removing unwanted air or bone that may be within the region of interest (135). For Study 1 the default of -1000 to 3000 was used; however, a more tailored approach was taken for the subsequent studies in which the intensity range was based on the minimum and maximum range of HUs present in the patients’ parotid glands. Specific details regarding pre-processing of each study are included in Table 2.3

Experimental Study	Region of Interest	Radiomics Software (Name, Version)	Voxel Resampling	Intensity Range	Discretization (Bin width)	No. of Shape Features	No. of Histogram Features	No. of Texture Features	Features Normalised?
1	Right PG	LIFEx v5.1	Resampling was not performed. In plane sampling = 0.97 x 0.97 mm/pixel. Slice thickness = 3mm	-1000 to 3000 HU	10 HU	4	6	32	No
2	Left PG	LIFEx v6.2	Naïve voxel size as well as resampled to 1x1x1mm were analysed	Cohort 1: -1000 to 1800 HU Cohort 2: -561 to 1632 HU	5 HU and 10 HU	5	6	32	No

3	Left and right PGs*	LIFEx v7.3	Resampled to 1x1x1mm voxel size	-1024 to 1739 HU	5 HU and 10 HU	14	30	56	Yes: z Score normalisation
4	Left and right PGs*	LIFEx v6.2	Resampled to 1x1x1mm voxel size	CT1: -569 to 1632 HU CT2 and CT3: -1000 to 1968 HU	5 HU	5	6	32	No

Table 2.3: Summary of radiomics pre-processing and feature extraction process. *RF values averaged between both glands. PG = parotid gland; HU = Hounsfield unit; CT = Computer tomography.

Following these pre-processing steps, radiomic feature extraction was carried out. The number of radiomic features is dependent on the version of the LIFEx software used. Whilst specific features extracted may vary between software versions, what is common are the feature families, which include shape (renamed morphological in LIFEx v7.3), (intensity) histogram, and texture features. In accordance with the IBSI reporting recommendations (125), details of the image acquisition and processing parameters are presented in Table 2.3, and the specific radiomic features extracted per study based on the version of LIFEx are detailed in Tables 2.1 and 2.2.

2.3 STUDY OUTCOMES

Study 1 examined the impact of interobserver contouring variability on radiomic feature values. This contouring variability was assessed using the Dice Similarity Coefficient and the Hausdorff Distance. Study 2 explored the impact of spatial resampling and intensity discretisation. For this study, the extracted radiomic feature values themselves were the study outcomes. Study 3 analysed clinical, dosimetric and radiomics data to predict xerostomia and sticky saliva. These toxicity outcomes were collected in line with the Randomized Trial of Avelumab-cetuximab-radiotherapy Versus Standard of Care in Locally Advanced Squamous Cell Carcinoma of the Head and Neck (REACH) which patients were recruited to (30), and utilised the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Head and Neck Module (EORTC QLQ-HN35) questionnaire (143). Data were collected at baseline, acute, sub-acute and late timepoints, guided by the trial protocol. Study 4 also examined acute and late xerostomia; however, in this study, the Radiation Therapy Oncology Group (RTOG) Acute Toxicity scale was used (73).

2.4 STATISTICAL ANALYSES

All data analysis was conducted with either Microsoft Excel or IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp.

Prior to statistical analysis, data were tested for normality using either the Jarque Berra or Shapiro Wilks test, informing the choice to use nonparametric statistical tests where appropriate. In addition to testing for normality, full assumption testing was carried out prior to conducting statistical analysis. Statistical significance was assumed based on a p value of <0.05 , except in the case of Study 4 which incorporated a Bonferroni correction to account for multiple testing, resulting in statistical significance of the Wilcoxon signed rank test being set at $p<0.017$.

Studies 1 and 2 considered the variation in radiomic feature values that may occur as a result of observer contouring variability, image resampling, or discretisation parameters. The magnitude of this potential dissimilarity in radiomic feature values was assessed using the paired percentage difference in Study 1 and the percentage coefficient of variance in Study 2.

Studies 3 and 4 incorporated clinical outcomes in which patient data as well as radiomic features values were tested for their association with xerostomia, with sticky saliva being included as an additional clinical endpoint in Study 3. In these studies, radiomic features were extracted from both the left and right parotid glands separately, and then averaged to obtain a single value per feature per patient. Whilst this approach has been taken in other radiomic studies (144), the suitability of this method was tested statistically using the Wilcoxon signed rank test.

In Study 3, association tests were conducted using chi-squared tests for the categorical variables (e.g., sex) and the Mann Whitney U test for the continuous variables (e.g., radiomic features or patient age). Given the work in Study 3 also included building a logistic regression model to predict

xerostomia and sticky saliva using a combination of radiomic and clinical independent variables, additional assumption testing was required. This included collinearity diagnostics using Spearman correlation, with a cut-off of ± 0.75 and variance inflation factor with a cut-off of 5. Feature reduction was performed using principal component analysis, in which case components were retained based on an Eigenvalue of >1 . The overall performance of the models was considered based on the balanced accuracy as well as the 2 log-likelihood and Nagelkerke R^2 . The discriminatory performance of the logistic regression models was measured using the area under the receiver operating characteristic (AUC-ROC) curve.

As mentioned, Study 4 examined the association of radiomic feature values with xerostomia; however, in contrast to Study 3, it was not the feature value extracted at baseline that was considered, but rather the *change* in radiomic feature values over the course of treatment. Radiomic features were extracted at mid treatment, “near end” of treatment, and 3-6 months following treatment completion. The overall variation in feature values during treatment was examined using the Friedmans test, with the Wilcoxon signed rank test used for post hoc analysis, including Bonferroni correction. The absolute paired percentage difference of each radiomic feature between the three scan timepoints was calculated and the association of this variability with xerostomia was tested using the Mann Whitney U test. A summary of all statistical tests used per study is provided in Table 2.4.

Experimental Study	Source of Data	Ethical Approval	Patient Characteristics		Outcomes and Data Analysis		
			No. of Patients	Site within the Head and Neck	Endpoints Considered	Statistical Tests Conducted	Data Analysis Software
1	ESTRO	TCD SOMREC	1	Oropharynx	Dice Similarity Coefficient; Hausdorff Distance; RF from planning CT	Paired Percentage Difference between radiomic features values	SPSS v26
2	TCIA	N/A; data is open source	Cohort 1 = 127 Cohort 2 = 17	Mixed	RF features from pre-treatment CT	Percentage Coefficient of Variance among radiomic features values	Microsoft Excel v16.66.1

3	Gustave Roussy	TCD SOMREC and Gustave Roussy	98	Mixed	Acute and late xerostomia and sticky saliva measured with EORTC HN-35 QoL; RF from planning CT	Chi squared, Mann Whitney U, Wilcoxon, Spearman, VIF, PCA, Logistic regression, 2 log-likelihood and Nagelkerke R2, AUC- ROC	SPSS v26
4	TCIA	N/A; data is open source	26	Mixed	Xerostomia measured with RTOG Acute Toxicity Scale; RF from CT at baseline, mid treatment and end of treatment	Freidman, Wilcoxon, Mann Whitney U	SPSS v26

Table 2.4: Summary of data source, patient cohort and data analysis. ESTRO = European Society for Radiotherapy and Oncology; TCIA = The Cancer Imaging Archive; TCD SOMREC = Trinity College Dublin School of Medicine Research Ethics Committee; EORTC HN-35 QoL = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Head and Neck Module; RTOG = Radiation Therapy Oncology Group; VIF = Variance Inflation Factor; PCA = Principal Component Analysis; SPSS = Statistical Package for the Social Sciences.

3.1 INTRODUCTION

Inter-observer contouring variability is a well-known challenge in radiation therapy, particularly in relation to the complex anatomy of head and neck region. Significant variability in both target and organ at risk delineation has been widely reported (145-151). Despite the aid of contouring guidelines, distinguishing the parotid gland boundaries from surrounding soft tissue remains challenging. The integration of magnetic resonance imaging has assisted in this regard; however, the impact of this approach on decreasing observer variability has shown mixed results (151, 152). This persistent variability impacts on dosimetric and clinical outcomes (151, 153-155), and recent studies have suggested the potential impact on computational image analysis (156-158). The quality of the delineation of the ROI will likely influence the value of extracted radiomic features, and some features may prove more robust than others against delineation uncertainties caused by multiple observers.

The impact of delineation uncertainties on radiomics feature value variability has been reported (157, 159-163). These studies have covered a range of tumour sites, including head and neck; however, radiomics analysis has been limited to target delineation, and variability in organ at risk contouring has not yet been investigated. Many of these studies are limited by the relatively small number of observers, relying on delineations of just up to five expert oncologists or radiologists. Given the range of radiation oncology professionals responsible for delineating organs at risk, this approach does not reflect practice in a true clinical environment.

This study will test the hypothesis that some radiomics features are more robust against inter-observer delineation variability than others, and as such may be more appropriate for use in the clinical environment. In order to identify these robust features, this study aimed to quantify the variability in the values of the radiomics features extracted from the whole right parotid as the region of interest delineated by a series of independent observers.

3.2 MATERIALS AND METHODS

3.2.1 Patient Data and Cohort of Participants

The whole right parotid of an anonymous patient with head and neck cancer was delineated on a 3D computer tomography (CT) image by 40 participants of an online workshop to generate a series of ROIs. The anonymous DICOM data generated from the workshop were transferred to the research team for subsequent radiomics analysis with institutional ethical approval. The SmartSegmentation tool in the Varian Eclipse treatment planning system (version 15.5) was used to generate an atlas-based auto-contour for comparison.

3.2.2 Quantifying Inter-Observer Delineation Variability

Inter-observer variability of the ROIs was quantified using the DICE similarity coefficient (DSC) for each participant's contour and the auto-segmentation result of the whole right parotid gland (164). The whole right parotid gland contour created by the ESTRO workshop's instructor was used as the reference (expert) contour (142). By considering the volume of overlap, the DSC quantifies the agreement of the participants' contours with that of the reference contour. A DSC of 0 indicates no overlap of contours (poor agreement) and a score of 1 indicates complete overlap (good agreement).

$$\text{Dice similarity} = 2 \times [X \cap Y] / [|X| + |Y|]$$

Where X is the reference contour, Y is the observer contour, and $[X \cap Y]$ is the volume overlap between the two contours.

In addition, the Hausdorff distance (HD) was also calculated using 3D Slicer software between the expert's contour and each participant, as well as the auto-contour (165). The HD more accurately detects variations in the ROIs shape that may not significantly alter the total volume (157). This metric is the maximum distance of each voxel of the participant's contour to the nearest point in the expert's contour. Lower values (measured in mm) indicate a higher agreement between the compared contours. The increased sensitivity of the HD in detecting deviations in contour shape that do not alter the ROI total volume makes it potentially more meaningful for radiomics analysis; especially considering Shape features extracted. To demonstrate the variation amongst the participant group, the coefficient of variation of the right parotid volume (in cc) was also calculated (standard deviation/mean), where a value as low as possible is desirable (151, 154).

Finally, with respect to observer delineation performance, outliers were identified based on the interquartile range method using SPSS (v26). For the DSC, outliers were those with a DSC less than $(Q1 - 1.5IQR)$, where Q1 is the first quartile and IQR the interquartile range. Conversely for HD, outliers were those with a HD greater than $(Q3 + 1.5IQR)$, where Q3 is the third quartile. To be classified as an outlier for this study, the participant had to meet both of these criteria.

3.2.3 CT Image and Radiomics Feature Extraction

Radiomics analysis was conducted using LIFEx software, version 5.10 (123), which complies with the Imaging Biomarker Standardisation Initiative recommendations (125). The in-plane sampling of the CT image was 0.97 x 0.97 mm/pixel and the slice thickness was equal to 3mm; these values from the native CT dataset were left unchanged. With respect to intensity rescaling, absolute resampling was used with a minimum bound of -1000

and a maximum bound of 3000. To reduce impact of noise, the number of grey levels was set to 400 with a bin size of 10 (166). These values are the default settings of LIFEx and were left unchanged with the view to investigate the impact of manipulating these specific settings in the subsequent chapter, Study 2.

The feature extraction was run to report on shape, first order features (histogram), as well as texture features. Shape features include sphericity, compacity, volume (mL) and volume (number of voxels). First order features derived from histograms include: skewness, kurtosis, excess kurtosis, entropy log10, entropy log2 and energy. Thirty two texture features from four main texture feature “families” were also extracted, as detailed in Chapter 2, Table 2.2.

3.2.4 Testing Robustness of Radiomics Features

The absolute paired percentage difference between each participant and the expert’s values for each radiomics feature was calculated:

$$\% \text{ Difference} = 100 \times | (RF_p - RF_e) / RF_e |$$

where RF_p is the radiomics feature value of the participant, and RF_e is that of the expert. Robustness of the features was defined using the classification approach described by Chou et al., (167). The features with mean absolute paired percentage difference <5% were considered robust. The features with mean absolute paired percentage difference >5% but <10% were considered stable. The features with mean absolute paired percentage difference >10% but <20% were considered acceptable. The features with mean absolute paired percentage difference >20% but <50% were classed as uncertain and all remaining features were considered unstable. All features were also ranked in ascending order by mean absolute paired percentage difference. As an additional analysis, the mean absolute paired percentage difference testing was repeated using the atlas-based auto-segmentation contour as a reference.

3.3 RESULTS

3.3.1 Inter-Observer Variability in the Generation of Right Parotid ROIs

A total of 40 participants from the online workshop were included in this study. The mean DSC was 0.68 ± 0.11 (median: 0.71; range: 0.32-0.78), with 63% of the participant group achieving a DSC >0.7 . The average HD for all the participants was $16.16\text{mm} \pm 4.16\text{mm}$ (median: 15.1mm; range: 10.12-28.95mm). The coefficient of variation of the participants' contours as a group was 0.16. Two observers were identified as outliers with a DSC of 0.32 and 0.45 as well as both having a HD of 28.95mm. The auto-segmentation approach achieved a DSC of 0.76 and a HD of 15.16mm.

Figure 3.1 shows the contours from the auto-segmentation approach along with a selection of three observers, with the expert contour as a reference. For participant #13 (Figure 3.1b) with a DSC of 0.5, 26% of the features were considered robust with an absolute paired percentage difference $<5\%$. In contrast, for participants #1 (Figure 3.1c) and #2 (Figure 3.1d) with DSC of 0.72 and 0.73 respectively, over half of the features were classed as robust. Similarly, for the auto-segmented result (Figure 3.1a) with a DSC of 0.76, 48% of features had an absolute paired percentage difference $<5\%$.

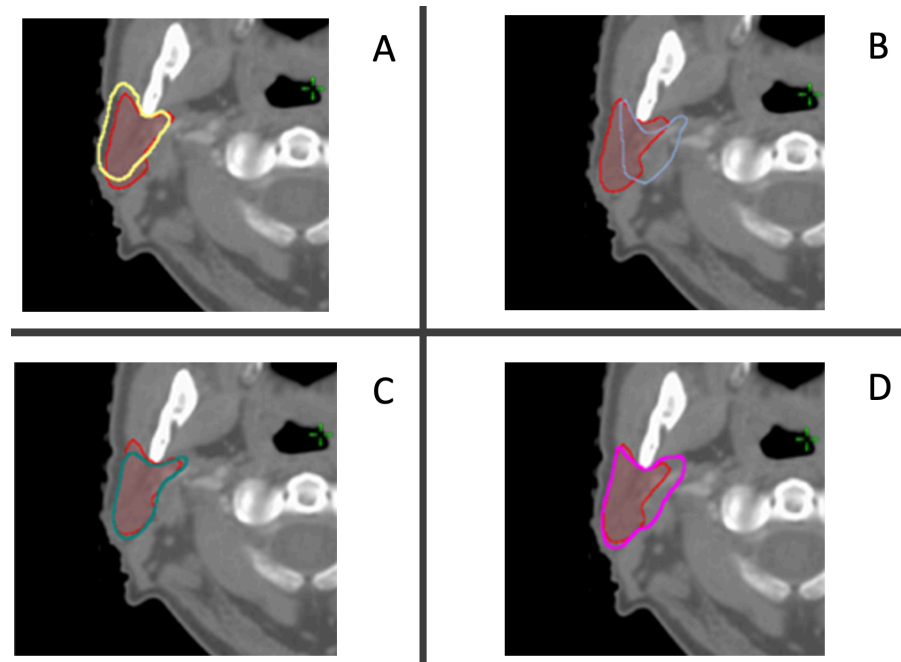


Figure 3.1: Variation in parotid gland contours. The contours of the right parotid gland defined by the expert is shown in red in each image above. The auto-segmentation contour (A) in yellow, DSC=0.76 when compared against the expert contour; Participant #13 contour (B) in violet, DSC=0.5 when compared against the expert contour; Participant #1 contour (C) in green, DSC=0.72 when compared against the expert contour; and Participant #2 contour (D) in pink, DSC=0.73 when compared against the expert contour.

3.3.2 Radiomics Features Values Variability Across Right Parotid ROIs

The mean, along with the standard deviation for the absolute paired percentage differences of radiomics features for the participant population are shown in Table 3.1, grouped by feature family. 33% (n=14) of the radiomics features were deemed to be robust with a mean absolute paired percentage difference <5%. Nine of these 14 features were part of the grey-level run length matrix family. Ten percent (n=4) of the features were deemed to be unstable with a mean absolute paired percentage difference

>50%, reaching a maximum mean absolute paired percentage difference of 65% for the Excess Kurtosis feature from the Histogram family.

The results of the additional testing, whereby the atlas-based auto-segmentation contour was used as the reference, along with the corresponding feature classification, are also shown in Table 3.1 as a comparison. The classification of 69% (n=29) of the features remained the same regardless of if the “expert” or auto-segmentation was used as the reference contour. Importantly, all features that were classed as “unstable” were consistent across both methods tested.

Table 3.2 presents the mean, along with the standard deviation, for the absolute paired percentage differences of radiomic features for the participant population with the two outlying observers removed. The classification of 86% (n=36) of the features was unaffected by the removal of these outliers.

Feature Family	Feature Name	AS Delineation Method Compared Using Expert as Reference	Manual Delineation Method Compared Using Expert as Reference			Manual Delineation Method Compared Using Auto-Segmentation as Reference			Consistency in Feature Classification*
			Mean	SD	Stability Classification	Mean	SD	Stability Classification	
Histogram	Skewness	70.88%	54.52%	42.63%	Unstable	238.47%	172.94%	Unstable	Y
	Kurtosis	68.64%	63.06%	50.18%	Unstable	269.70%	234.79%	Unstable	Y
	ExcessKurtosis	70.41%	64.69%	51.47%	Unstable	293.19%	255.24%	Unstable	Y
	Entropy_log10	6.62%	5.40%	5.45%	Stable	8.19%	6.46%	Stable	Y
	Entropy_log2	6.62%	5.40%	5.45%	Stable	8.19%	6.46%	Stable	Y
	Energy	16.82%	12.50%	14.87%	Acceptable	24.62%	22.14%	Uncertain	N
Shape	Volume (mL)	16.38%	25.28%	17.78%	Uncertain	19.76%	18.53%	Acceptable	N
	Volume (# vx)	16.38%	25.28%	17.78%	Uncertain	19.76%	18.53%	Acceptable	N
	Sphericity	10.70%	4.75%	3.06%	Robust	6.74%	4.21%	Stable	N
	Compacity	11.78%	14.34%	7.62%	Acceptable	9.09%	10.97%	Stable	N
Grey level co-occurrence matrix (GLCM)	Homogeneity	4.00%	4.06%	4.28%	Robust	6.24%	5.21%	Stable	N
	Energy	24.90%	22.92%	30.72%	Uncertain	43.75%	48.27%	Uncertain	Y
	Contrast	22.48%	58.15%	75.61%	Unstable	68.32%	107.36%	Unstable	Y
	Correlation	32.94%	20.52%	13.37%	Uncertain	12.51%	8.22%	Acceptable	N

	Entropy_log10	6.37%	5.07%	5.56%	Stable	7.82%	6.43%	Stable	Y
	Entropy_log2	6.37%	5.07%	5.56%	Stable	7.82%	6.43%	Stable	Y
	Dissimilarity	4.84%	14.52%	13.22%	Acceptable	15.64%	12.73%	Acceptable	Y
Grey-level run length matrix (GLRLM)	SRE	0.87%	1.13%	1.09%	Robust	1.43%	1.16%	Robust	Y
	LRE	3.84%	4.34%	4.33%	Robust	5.89%	4.97%	Stable	N
	LGRE	2.08%	0.68%	0.50%	Robust	2.63%	0.59%	Robust	Y
	HGRE	2.00%	0.94%	1.39%	Robust	2.89%	1.49%	Robust	Y
	SRLGE	3.15%	1.42%	1.09%	Robust	3.83%	1.56%	Robust	Y
	SRHGE	1.35%	1.92%	2.03%	Robust	2.59%	2.33%	Robust	Y
	LRLGE	2.47%	4.50%	4.45%	Robust	4.90%	4.91%	Robust	Y
	LRHGE	5.09%	3.94%	4.18%	Robust	7.21%	4.84%	Stable	N
	GLNU	1.23%	20.94%	13.39%	Uncertain	21.71%	13.47%	Uncertain	Y
	RLNU	20.08%	27.86%	18.64%	Uncertain	20.67%	20.32%	Uncertain	Y
	RP	1.31%	1.50%	1.47%	Robust	1.92%	1.53%	Robust	Y
Neighbourhood grey-level different matrix (NGLDM)	Coarseness	2.70%	35.24%	55.13%	Uncertain	35.55%	57.52%	Uncertain	Y
	Contrast	8.85%	41.62%	52.06%	Uncertain	35.34%	46.42%	Uncertain	Y
	Busyness	17.89%	45.04%	28.99%	Uncertain	33.89%	23.28%	Uncertain	Y
	SZE	7.99%	4.47%	3.22%	Robust	4.97%	3.01%	Robust	Y
	LZE	4.27%	25.48%	15.60%	Uncertain	28.37%	15.99%	Uncertain	Y

Grey-level zone length matrix (GLZLM)	LGZE	4.22%	1.67%	1.79%	Robust	5.46%	1.88%	Stable	N
	HGZE	4.72%	3.61%	6.38%	Robust	7.77%	7.17%	Stable	N
	SZLGE	4.07%	5.88%	3.37%	Stable	2.90%	2.79%	Robust	N
	SZHGE	12.89%	8.39%	8.74%	Stable	14.23%	13.61%	Acceptable	N
	LZLGE	4.46%	25.55%	15.61%	Uncertain	28.57%	16.04%	Uncertain	Y
	LZHGE	4.08%	25.41%	15.59%	Uncertain	28.15%	15.95%	Uncertain	Y
	GLNU	15.39%	24.39%	18.34%	Uncertain	21.00%	18.10%	Uncertain	Y
	ZLNU	4.90%	31.19%	21.03%	Uncertain	33.41%	20.95%	Uncertain	Y
	ZP	4.10%	13.11%	8.50%	Acceptable	12.11%	9.19%	Acceptable	Y

Table 3.1: Absolute paired percentage difference of each radiomic feature, along with the standard deviation, and stability classification. Data presented demonstrates the absolute paired percentage differences when the auto-segmentation results are compared against the expert; the mean absolute paired percentage differences when the observers are compared against the expert as a reference; the mean absolute paired percentage differences when the observers are compared against the auto-segmentation as a reference. *Consistency in classification refers to a comparison between the AS being used as the reference and the expert being used as the reference. AS=Auto-segmentation; SD=Standard deviation.

Feature Family	Feature Name	All observers (n=40)			After Removal of Outliers (n=38)			Consistency in Feature Classification
		Mean	SD	Stability Classification	Mean	SD	Stability Classification	
Histogram	Skewness	54.52%	42.63%	Unstable	52.43%	42.72%	Unstable	Y
	Kurtosis	63.06%	50.18%	Unstable	61.34%	50.92%	Unstable	Y
	ExcessKurtosis	64.69%	51.47%	Unstable	62.92%	52.23%	Unstable	Y
	Entropy_log10	5.40%	5.45%	Stable	4.77%	4.49%	Robust	N
	Entropy_log2	5.40%	5.45%	Stable	4.77%	4.49%	Robust	N
	Energy	12.50%	14.87%	Acceptable	10.83%	12.07%	Acceptable	Y
Shape	Volume (mL)	25.28%	17.78%	Uncertain	23.00%	14.90%	Uncertain	Y
	Volume (# vx)	25.28%	17.78%	Uncertain	23.00%	14.90%	Uncertain	Y

	Sphericity	4.75%	3.06%	Robust	4.76%	3.13%	Robust	Y
	Compacity	14.34%	7.62%	Acceptable	13.46%	6.61%	Acceptable	Y
Grey level co-occurrence matrix (GLCM)	Homogeneity	4.06%	4.28%	Robust	3.57%	3.48%	Robust	Y
	Energy	22.92%	30.72%	Uncertain	19.13%	22.77%	Acceptable	N
	Contrast	58.15%	75.61%	Unstable	57.11%	77.45%	Unstable	Y
	Correlation	20.52%	13.37%	Uncertain	20.61%	13.70%	Uncertain	Y
	Entropy_log10	5.07%	5.56%	Stable	4.41%	4.43%	Robust	N
	Entropy_log2	5.07%	5.56%	Stable	4.41%	4.43%	Robust	N
	Dissimilarity	14.52%	13.22%	Acceptable	13.40%	12.25%	Acceptable	Y
Grey-level run length matrix (GLRLM)	SRE	1.13%	1.09%	Robust	1.06%	1.07%	Robust	Y
	LRE	4.34%	4.33%	Robust	4.08%	4.19%	Robust	Y
	LGRE	0.68%	0.50%	Robust	0.70%	0.50%	Robust	Y
	HGRE	0.94%	1.39%	Robust	0.97%	1.42%	Robust	Y
	SRLGE	1.42%	1.09%	Robust	1.35%	1.07%	Robust	Y

	SRHGE	1.92%	2.03%	Robust	1.89%	2.07%	Robust	Y
	LRLGE	4.50%	4.45%	Robust	4.25%	4.32%	Robust	Y
	LRHGE	3.94%	4.18%	Robust	3.69%	4.04%	Robust	Y
	GLNU	20.94%	13.39%	Uncertain	19.04%	10.60%	Acceptable	N
	RLNU	27.86%	18.64%	Uncertain	25.62%	16.09%	Uncertain	Y
	RP	1.50%	1.47%	Robust	1.44%	1.47%	Robust	Y
Neighbourhood grey-level different matrix (NGLDM)	Coarseness	35.24%	55.13%	Uncertain	25.14%	27.32%	Uncertain	Y
	Contrast	41.62%	52.06%	Uncertain	41.44%	53.45%	Uncertain	Y
	Busyness	45.04%	28.99%	Uncertain	45.57%	29.54%	Uncertain	Y
Grey-level zone length matrix (GLZLM)	SZE	4.47%	3.22%	Robust	4.12%	2.71%	Robust	Y
	LZE	25.48%	15.60%	Uncertain	24.54%	15.38%	Uncertain	Y
	LGZE	1.67%	1.79%	Robust	1.68%	1.84%	Robust	Y
	HGZE	3.61%	6.38%	Robust	3.75%	6.52%	Robust	Y

	SZLGE	5.88%	3.37%	Stable	5.57%	3.00%	Stable	Y
	SZHGE	8.39%	8.74%	Stable	8.10%	8.82%	Stable	Y
	LZLGE	25.55%	15.61%	Uncertain	24.63%	15.42%	Uncertain	Y
	LZHGE	25.41%	15.59%	Uncertain	24.45%	15.35%	Uncertain	Y
	GLNU	24.39%	18.34%	Uncertain	22.60%	16.99%	Uncertain	Y
	ZLNU	31.19%	21.03%	Uncertain	28.60%	18.11%	Uncertain	Y
	ZP	13.11%	8.50%	Acceptable	12.76%	8.58%	Acceptable	Y

Table 3.2: Mean absolute paired percentage difference of each radiomic feature, along with the standard deviation, and stability classification. Data presented demonstrates absolute paired percentage differences for the whole cohort (n=40) and with the two outliers removed (n=38). SD=Standard deviation.

3.3.3 Identification of the Most and Least Robust Radiomics Features

Figure 3.2 shows the tabulated values from Table 3.1, where the expert's contour has been used as the reference contour; the results have been ranked in ascending order of the magnitude of the mean absolute paired percentage difference.

Taking the two radiomics features with the highest and lowest mean absolute paired percentage differences, Figures 3.3a and 3.3b display the relationship between the absolute paired percentage differences and each participant's DSC and HD scores respectively; demonstrating that for the most robust feature (GLRLM_LRGE), the radiomics feature value does not vary irrespective of DSC value or HD. Conversely, the least stable feature (HISTO_ExcessKurtosis) shows greater variability amongst the participant group, where even for the participant with the highest DSC of 0.78 and a HD of 10.12mm, the absolute paired percentage variation was still as high as 64%. Figure 3.4 shows similar results for these two specific features when the absolute paired percentage difference in radiomics features values are plotted against the absolute paired percentage difference in the size of the ROI, measured by the total number of voxels. For the least robust feature (HISTO_ExcessKurtosis), a percentage difference in number of voxels of 7% between an individual observer and the expert resulted in a difference in radiomics features values of 165%. In contrast, the GLRLM_LGRE feature value remained within 1% of the expert's result despite the number of voxels within the ROI differing by up to 78%.

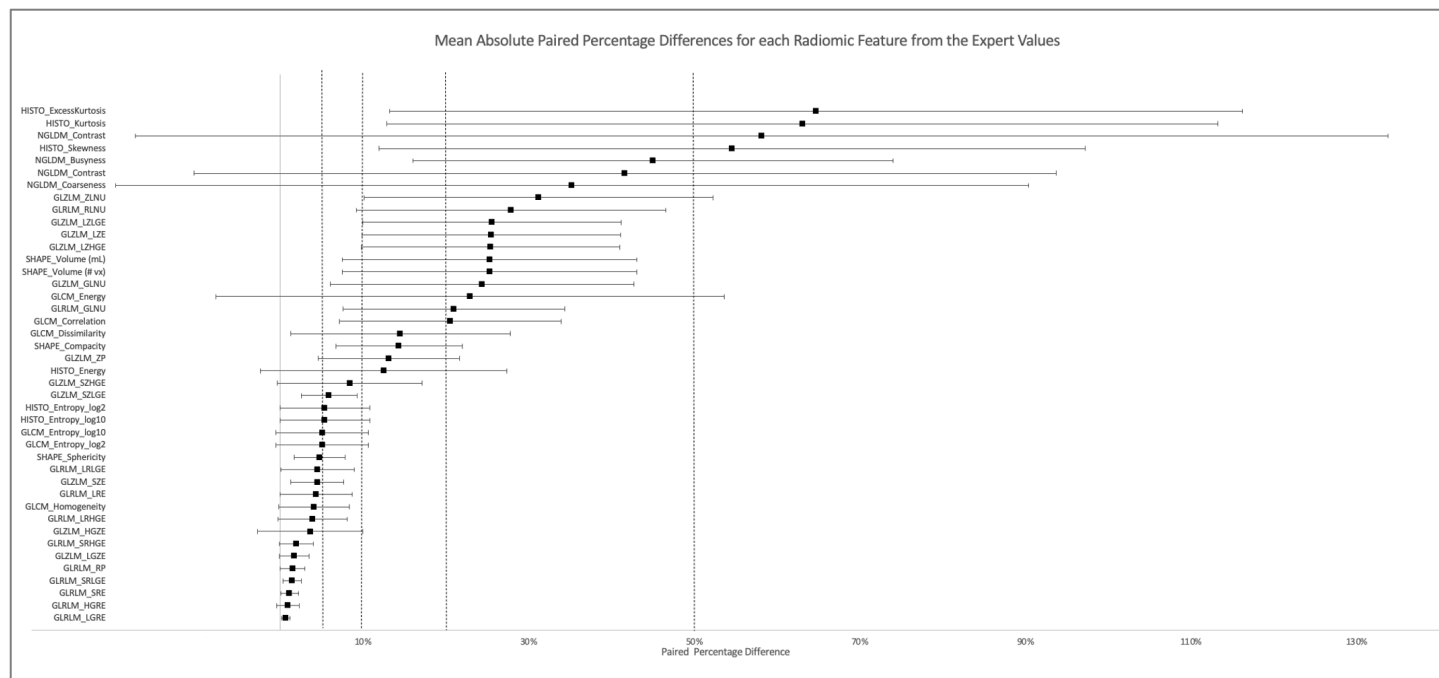


Figure 3.2: Plot of the mean absolute paired percentage differences of the participants' values from the expert values ranked according to increasing mean. Error bars illustrate the standard deviation. Dotted vertical lines demarcate the mean absolute paired percentage difference thresholds used to classify feature robustness.

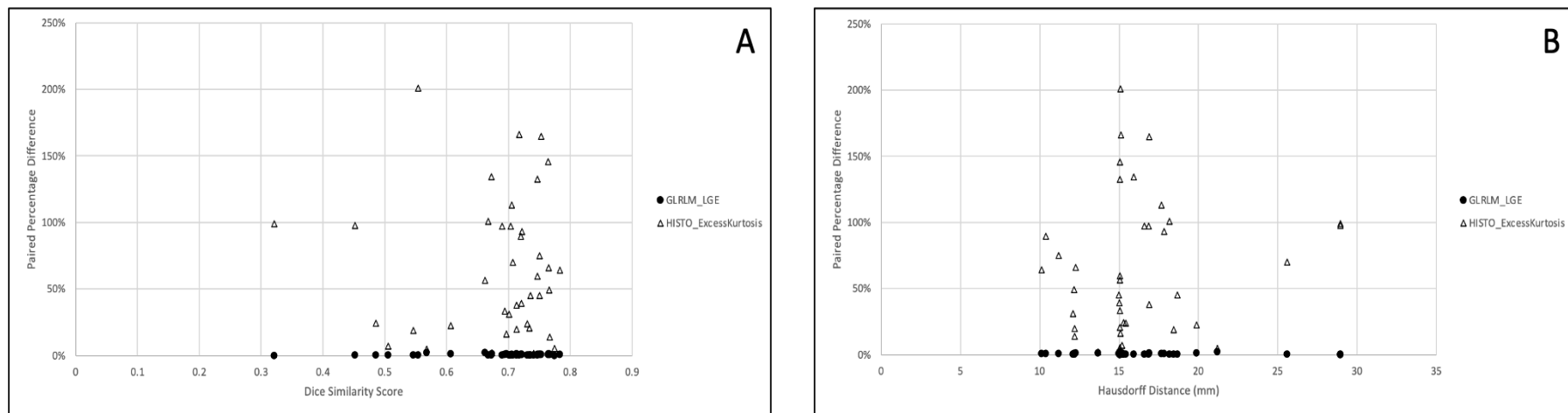


Figure 3.3: Scatter plot of the absolute paired percentage differences in feature value between participant and expert against their corresponding DICE similarity coefficient (A) and Hausdorff distance (B). Data related to the most and least robust features identified as those with the smallest (GLRML_LGRE) and the largest (HISTO_ExcessKurtosis) mean absolute paired percentage difference is represented.

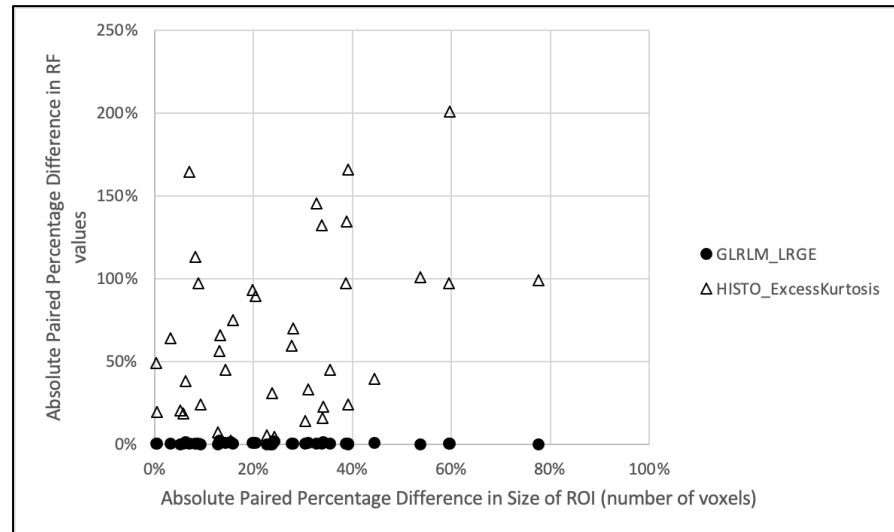


Figure 3.4: Scatterplot for the least and most robust features, based on the mean absolute paired percentage difference. Paired percentage differences in radiomics feature values with the expert values plotted against the absolute paired percentage differences in the size of the region of interest, measured in number of voxels. ROI = region of interest; RF = radiomics feature

3.4 DISCUSSION

Inter-observer delineation variability is problematic in radiation therapy, influencing the quality of treatment planning (151). This variability also affects the identification of radiomics features relevant to the prediction of presence of specific clinical endpoints. This study specifically focused on the evaluation of the sensitivity of radiomic features values to the variability of user-defined and computer-generated ROIs of the right parotid in a single test Head and Neck patient. Although the mean DSC for this cohort of 40 participants was 0.68 (± 0.11), it must be noted that 63% of the participants achieved a DSC > 0.7 ; a threshold considered in the literature to indicate acceptable delineation agreement with the expert (154). Individual

participant's clinical experience in OAR contouring could explain some of the DSC and HD results achieved here in this study; however, without this information available it is not possible to draw any conclusions in this regard.

Also using the DSC, several studies have quantified inter-observer variability specifically for the parotid glands. Based on a cohort of seven radiation oncology residents and working off a single patient case, Awan et al., (168) reported a mean DSC of 0.85 (± 0.06) following the introduction of contouring guidelines. Using a larger number of cases ($n=40$), the cohort from Walker et al., (169) achieved a similar mean DSC of 0.90 (± 0.11). These results, albeit impressive, are limited by the small cohort size of just 7 and 8 observers respectively. A larger study by Nelms et al., (151) of 32 observers achieved a DSC of 0.77 (± 0.08) and a coefficient of variation of 0.29; slightly worse than the participant group of this study who achieved a coefficient of variation of 0.16.

In addition to the established observer delineation variability, contouring is also recognised to be one of the most time-consuming tasks in radiation therapy. In an attempt to overcome this, many clinical departments are utilising auto-segmentation software to expedite this process. Although the DSC based on auto-segmentation method was 0.76 (0.08 higher than this cohort's average) it was still not as high as the auto-segmentation result reported by Walker et al., (169) of 0.89 (± 0.20), or that reported by the performance from the deep learning methods used by Hänsch et al., (170) (DSC of 0.88).

A number of studies have investigated the impact of delineation variability on radiomics features, with the majority of research focusing on lung tumours using the target volume as the region of interest (157, 158, 163, 171). Owens et al., (157) were the only ones to quantify the agreement of auto-segmented ROIs with that of expert consensus prior to feature analysis (DSC = 0.88). Classifying robustness of 40 features according to the intra-class correlation coefficient (ICC), the authors were unable to conclude that any one category of features was more robust than another.

These results are in contrast to Parmar et al., (158) who highlighted high reproducibility of intensity and texture features following 3D-Slicer based segmentations when compared to manual delineations. Furthermore, using an ICC cut off of 0.8, they reported that 70% of features extracted based on auto-segmentation were more stable than those of manual delineation methods (52%). Using area under the curve (AUC) analysis, Haga et al., (163) also demonstrated that inter-observer variability significantly impacts on radiomics performance, and they went on to definitively state that LHH_Skewness, HLH_RLV and, HHL_LRHGE wavelet features are both clinically meaningful for the prediction of histology of non-small cell lung cancers, and also robust to observer delineation variability.

Pavic et al., (161) also demonstrated that DSC is correlated with radiomics features values; where the disease site with the worst DSCs, i.e. mesothelioma, also had the least number of stable features (36%). Conversely the lung cases had the least observer variability, reflected in a mean DSC of 0.89; and consequently, had the highest number of stable features (90%). Conversely Mori et al., (159) reported inter-observer variability to have a minimal impact; with just 4 out of 69 features (5%) to have an ICC<0.8. However, arguably there was limited scope to investigate this effect given the acceptable DSCs achieved by the three observers (DSC = 0.76 ± 0.09). Although the range in DSC by Pavic et al., (161) for the mesothelioma cases (DSC = 0.0-0.9) and head and neck cases (DSC = 0.21-0.89) could have provided scope for more detailed analysis, their study was also limited by small sample size of just three observers. Pavic et al., (161) also cite the single institutional nature of their study as a limiting factor and call for the publication of similar studies in order to generate a collection of data across institutions. The large spread of DSCs from the cohort of Study 1 (DSC range: 0.32-0.78) along with the larger sample size of participants (n=40), allowed for feature robustness to be tested under a range of contouring conditions. Whilst two outliers were identified in this study, they were purposefully not removed from the main analysis for this reason. Whilst some features, namely HISTO_Entropy and GLCM_Entropy did show an improvement in stability classification moving from “Stable” to

“Robust”, the actual mean absolute paired percentage difference only improved by 0.63% and 0.66% respectively. Nonetheless closer inspection of the robustness of some features may be warranted; taking the specific observer contours presented in Figure 3.1, three features (GLRLM_RLNU, NGLDM_Bussyness, and GLZLM_GLNU) appeared robust to the contour variation of participant #13 (DSC = 0.5) with a paired percentage difference <5%. Despite acceptable DSC of 0.72 for participant #1, these features (GLRLM_RLNU, NGLDM_Bussyness, and GLZLM_GLNU) had an absolute paired percentage difference of 22%, 12% and 22% respectively when compared with the reference contour; and rose to 41%, 45% and 47% respectively for participant #2 (DSC = 0.73).

The use of DSC is common practice in radiation oncology delineation studies; however, it has been criticised due to the dependency of the size of the ROI in question, where a higher DSC is more easily achieved for larger volumes (172). For this reason, the relationship between the HD and radiomics feature values was also explored. The results in Figures 3.3a and 3.3b essentially demonstrate that whether the DSC or the HD is considered, the performance of the most and least robust features identified in this study behave in a similar fashion. In the case of the least robust feature (HISTO_ExcessKurtosis), when HD calculated for multiple observers was identical at 15mm, the paired percentage difference in the feature value calculated ranged from less than 10% for one observer to nearly 200% for another. Similarly, when the DSC varied by <0.1 between observers, the paired percentage difference of this feature ranged from less than 10% to over 150%.

In the context of Positron Emission Tomography (PET) imaging, the following texture features were considered robust to segmentation variation: GLCM_Homogeneity, GLCM_Entropy, GLRLM_Short-Run Emphasis, GLRLM_Long-Run Emphasis, GLZLM_Low Grey-level Zone Emphasis and, GLZLM_High Grey-level Zone Emphasis (133). These data presented here suggests that these features are also robust to observer delineation variability when using CT imaging, based on the parotid gland

identified as the ROI. To supplement these listed above, this study has also identified GLRLM_LGRE, GLRLM_HGRE, GLRLM_SRLGE, GLRLM_SRHGE, GLRLM_LRLGE, GLRLM_LRHGE, GLRLM_RP and, GLZLM_SZE to be robust with mean absolute paired percentage differences <5% when compared to either the expert's value or that of the auto-segmentation method. For this study, SHAPE_Sphericity was also robust with a mean absolute paired percentage difference of 4.75% and a standard deviation of 3.06% for the manually generated contours. This result was surprising given shape features for CT imaging under manual delineation conditions are thought to be less stable (137); however, the difference of 10.70% for the auto-segmentation method does cast doubt over conclusions drawn here. Also using atlas based auto-segmentation, Delaney et al., (173) reported an average percentage of region overlap between the physician's parotid contour and the auto-contour as low as $75.5\% \pm 12.0\%$. Whilst they demonstrated that the dosimetric impact of this may be negligible, when performing computational analysis at a voxel level, as is the case in radiomics, these differences become more pronounced.

With the exception of Mori et al., (159) many studies, including those examining radiomics features extracted from PET images (160, 171) have all concluded that observer delineation variability has an impact on the output of radiomics values, with some studies highlighting specific features or feature families to be more robust to this phenomenon. The results presented here agree with the consensus statements in the review by Traverso et al., (137). Considering a range of factors that could influence feature stability, they conclude that entropy derived from the histogram was the most stable first order feature. Results presented here confirm this feature's stability in which entropy had a mean absolute paired percentage difference of $5.40\% \pm 5.45\%$. Conversely, kurtosis taken from the histogram feature group was deemed unstable with a mean absolute paired percentage difference in excess of 50% for both the manually and auto-segmented contours; a result which again supports Traverso et al., (137). With respect to texture features, Traverso et al., (137) highlight a lack of clarity regarding isolating the most robust features. Although they do point

out that coarseness and contrast are among the least reproducible, it is not made clear which feature group they are referring to (i.e., GLCM or NGLDM). In this study, the contrast feature of the grey level co-occurrence matrix group was also considered to be a highly unstable feature with a mean absolute paired percentage difference of 58.15%; whilst this was not the largest mean absolute paired percentage difference, this feature did have the largest standard deviation of 75.61%, indicating variability amongst the observers' radiomic values.

This study is focused on the analysis of the variability of radiomics features within the ROIs of the right parotid of one patient across multiple observers. Whilst informative, this study does have limitations which must be presented. Firstly, when making a link between radiomic features and clinical outcomes inter-patient effects must also be considered. To do this confidently, radiomic studies typically involve images from a large number of patients to ensure observations are not patient specific. Both Haga et al., (163) and Huang et al., (162) used images from over 40 patients, however this benefit was at the expense of the number of observers completing the delineation, where both studies were limited to just three observers. For this study a large number of delineation observers was favoured over a large number of patient cases to answer the specific research question. Whilst the large number of observers is a strength of this study, the clinical experience of the cohort was unknown. Considering the data were taken from participants of an educational workshop, it could be postulated that their clinical experience in delineation is limited; however, this is pure speculation. Regardless, further evaluation of the identified features using intra-class correlation coefficient calculations as an indicator for feature robustness across many patient cases is warranted (174-176). Similarly, the single patient nature of this study meant just a single scanner was used and features were extracted at baseline only. In addition only the whole right parotid gland was analysed as the ROI, and although there was scope to also extract features from the left parotid gland, this was not done. These factors could potentially influence values of the extracted feature. Finally, the evaluation of a wider range of software solutions for quantitative

imaging analysis (e.g. OncoRadiomics, Pyradiomics, or programs developed in-house with Python) (162, 163) will influence repeatability and reproducibility and has the potential to identify alternative features (137); however these uncertainties may be mitigated by using IBSI compliant software.

The application of radiomics for the prediction of xerostomia continues to garner interest in the literature with several researchers identifying potentially informative features in this regard (116, 177, 178). Although the clinical relevance of these findings was outside the scope of this present study, it has demonstrated that robustness of features is varied, and careful selection is recommended when applying them in a clinical context.

3.5 CONCLUSION

Quantitative imaging analysis can be influenced by inter-observer delineation variability hence potentially limiting the reliability under clinical applications. The use of auto-segmentation may improve the concordance of feature results; however, this would need to be tested for a larger number of patient cases. Similarly, examining a larger patient cohort will determine if observations made in this single patient study are patient specific. This study did, however, isolate a number of radiomics features which appear robust to uncertainties related to delineation variability. These specific features now need to be investigated to ascertain their clinical value for the prediction of xerostomia.

4.1 INTRODUCTION

Computer tomography images, when taken for routine clinical use, can vary significantly in terms of acquisition protocols and reconstruction parameters; all of which impact on the image noise, contrast and spatial resolution and hence the radiomic feature values extracted. As a result, the subtle underlying biological characteristics of the patient may be muted, hindering analysis and clinical application. Several studies have investigated the impact of specific CT settings on radiomic features, and whilst they do use the more sophisticated Credence Cartridge Radiomics (CCR) phantom, they do not represent the diversity and complexity of patients' tissue (179-181). The few studies which also included a clinical cohort were limited to lung cancer patients (180, 182, 183), and more recently pancreatic cancer (159).

Recent publications indicate the potential for radiomics to predict xerostomia and structural changes in the parotid gland itself (177, 184, 185). Whilst these studies are novel in their application of radiomics, there are no studies that assess the impact of scan acquisition and/or reconstruction methods where the parotid gland is the region of interest. This has identified a gap in the literature; whereby given the complexity of the glands' cellular makeup, extrapolating conclusions based on other disease sites is not appropriate.

A naïve approach would be to simply confine radiomic analyses to patient populations in which the images were taken under identical conditions. In the era of big data and data mining, this is simply not pragmatic in the search for imaging biomarkers. There is a recognised need for continued

dedicated research in this field to establish appropriate harmonisation strategies when utilising large multi-institutional patient cohorts, such as those in clinical trials or imaging archives (128, 135).

The aim of this study was to quantify the impact of spatial resampling on radiomic features reproducibility associated with CT scanners and scan protocols heterogeneity within a multi-institutional cohort of patients with head and neck cancer. The reproducibility of radiomic feature values was investigated under different fixed bin sizes used for intensity discretization.

4.2 MATERIALS AND METHODS

4.2.1 Patient Dataset And Scan Parameters

Imaging data for this study were taken from the open access Cancer Imaging Archive (TCIA), and as such did not require institutional ethical approval for this secondary analysis (138). Given the primary aim of this study was to investigate the impact of spatial resampling when faced with variability in scanner protocol, the TCGA-HNSC dataset (Cohort 1) was chosen specifically as it states “the image data sets are also extremely heterogeneous in terms of scanner modalities, manufacturers and acquisition protocols” (139). From this dataset a total of 127 patients across eight different clinical departments were included in this study.

In addition to this group, a separate validation cohort (Cohort 2) was also taken from TCIA (138, 140). This cohort of 17 patients with head and neck cancer was selected as the patient images were all acquired under the same conditions, and all were taken for the purpose of radiotherapy treatment planning.

4.2.2 ROI Delineation

The whole left parotid gland was chosen as the ROI for this study. To limit inter-observer variability all delineations on Cohort 1 were completed by a single observer, following published guidelines (142). The initial contour was generated using atlas-based auto-contouring software within the Varian Eclipse treatment planning system (version 15.5). Following visual inspection, manual adjustments were made where the auto-contour generated failed to produce an acceptable contour. For the separate validation cohort (Cohort 2), the whole left parotid gland was already delineated as part of clinical care, and this contour was left unadjusted. Following this, the CT image and resultant RT Structure Sets were exported for radiomic analysis.

4.2.3 Feature Extraction

Radiomic analysis was conducted using the open access LIFEx software (version 6.2). For intensity rescaling of Cohort 1, absolute resampling was used with the minimum bound of -1000 and a maximum bound of 1800. These values were selected by first measuring the HU range within the left parotid ROI of the whole group. Likewise, this approach to parameter selection was also used for the Cohort 2 in which the HUs ranged from -561 to 1632.

With respect to grey level discretization, there are two approaches recognised by the Image Biomarker Standardisation Initiative (186); that is, fixing either the number of bins or the bin size. The latter approach was investigated in this study (125). Whilst there is no consensus regarding the most appropriate bin size for parotid radiomic studies (116), it is recognised that excessive reduction in the bin size would render the discretization process redundant. For Cohort 1 (n=127), the default bin size of 10 was initially used. The radiomics feature extraction was then repeated following adjustment to a fixed bin size of 5, resulting in 561 grey levels. Of note, the HU range was limited to a maximum value of 1800 when the bin size was

adjusted to 5, yet the limit applied under the bin size of 10 was the default setting of 3000.

As detailed in Chapter 2, (Tables 2.1 and 2.2), the feature extraction was run to report on shape features, including sphericity, compacity, volume (mL) and volume (number of voxels). First order histogram features extracted included skewness, kurtosis, excess kurtosis, entropy log10, entropy log2 and energy. In addition, conventional first order indices relating to Hounsfield Units within the ROI were recorded. Finally, 32 texture features were extracted.

4.2.4 Voxel Resampling

For Cohort 1, the in-plane sampling of the CT images ranged from 0.369-1.367 mm/pixel and the slice thickness ranged from 1 – 5mm. In an attempt to correct for this variation in voxel size and reduce the impact on radiomic feature values, spatial resampling to an isotropic voxel size of 1x1x1mm, using polynomial lagrangian interpolation within LIFEx was used. To assess the impact of this method of voxel resampling on radiomic feature values, feature extraction was completed on the native CT image and then again following spatial resampling. The difference in the mean values for each radiomic feature of the cohort following resampling was calculated. The difference in the standard deviation was also computed. Finally, the absolute value of the percentage coefficient of variation (%COV) following resampling was evaluated as:

$$\%COV = |(\sigma/\text{Mean}) \times 100|$$

The radiomic features were classified based on the influence spatial resampling had on the %COV for that feature. Using the same classification and threshold values as Shafiq-Ul-Hassan et al. (179), Class 1 were RFs where the %COV was <30%, Class 2 were RF where the %COV was >30% but <50%. Finally, Class 3 were RF whose %COV was >50%. This analysis and classification process was performed on both bin sizes in Cohort 1.

Finally, as a comparison the same spatial resampling and analysis methodologies were applied to Cohort 2, with a fixed bin size of 5. The above process is represented in Figure 4.1.

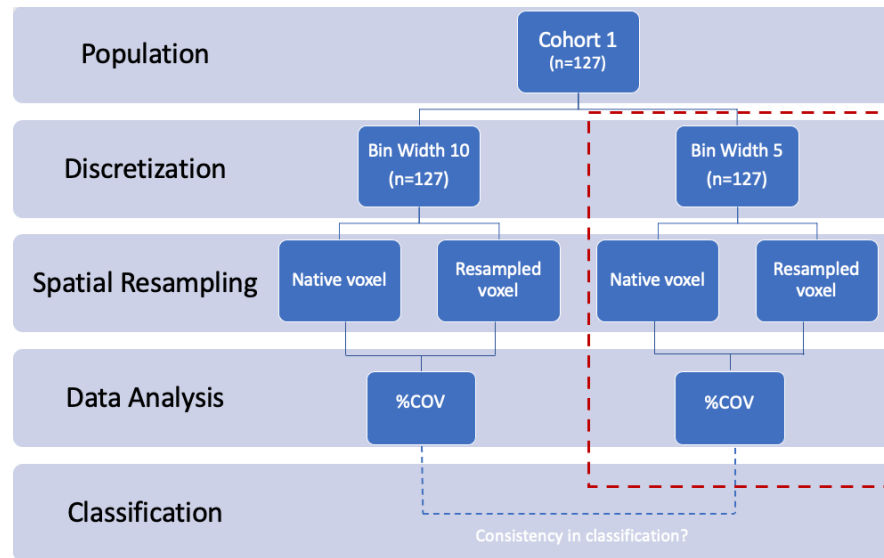


Figure 4.1: Process of methodology including intensity discretisation, spatial resampling, feature extraction, data analysis and classification. As a comparison, the same process contained within the red dotted lines were also applied to Cohort 2 (n=17). %COV = percentage coefficient of variance.

4.3 RESULTS

4.3.1 Cohort 1

For Cohort 1, a total of 127 patients across eight different departments were included. Relevant metadata extracted from the DICOM header can be found in Table 4.1. This data shows the significant heterogeneity in the image acquisition and reconstruction protocols used, with many parameters being adjusted on an individual patient basis. For the cohort overall, the pixel size ranged from 0.369-1.367 mm (mode = 0.488mm x 0.488mm, n=36), and the slice thickness ranged from 1 – 5mm (mode = 114

1.25mm, n=43). Each patient in Cohort 2 (n=17) had a common pixel size of 0.9765mm x 0.9765mm and a slice thickness of 2mm. All images in Cohort 2 were acquired on a Siemens Sensation Syngo CT 2009E; 120kVp, 240mA, 400mAs, with both a data collection and reconstruction diameter of 500 mm, using a B30s convolution kernel.

	Slice thickness (mm)	Pixel Spacing (mm)	Data Collection Diameter (mm)	Recon. Diameter (mm)	Recon. Kernel	Tube voltage (kVp)	Tube Current (mA)	Exposure (mAs)	Exposure Time (msec)
GE (n= 50)	1.25 (n=39)	0.449 x 0.449 (n=1)	250 (n=11)	230 (n=1)	Lung (n=1)	120 (n=31)	Range: 30 - 500	Range: 1 - 15620	Range: 500 - 1782
	2.50 (n=4)	0.473 x 0.473 (n=1)	320 (n=3)	242 (n=1)	Soft (n=3)	140 (n=19)	Median: 220	Median: 18	Median: 1000
	3.27 (n=1)	0.488 x 0.488 (n=36)	500 (n=36)	250 (n=36)	Standar d (n=46)		Mode: 220	Mode: 20	Mode: 1000
	3.75 (n=6)	0.523 x 0.523 (n=1)		268 (n=1)					

0.576 x							
0.576 (n=1)		295 (n=1)					
0.611 x							
0.611 (n=1)		313 (n=1)					
0.625 x							
0.625 (n=1)		320 (n=1)					
0.781 x							
0.781 (n=1)		400 (n=1)					
0.977 x							
0.977 (n=5)		500 (n=5)					

	1.367 x 1.367 (n=1)		700 (n=2)					

TOSHIBA (n=24)	2.00 (n=21)	0.415 x 0.415 (n=1)	240 (n=2)	213 (n=1)	FC04 (n=6)	120 (n=24)	Range: 150 - 300	Range: 75 - 150	500 (n=23)
	3.00 (n=3)	0.429 x 0.429 (n=10)	320(n=21)	220 (n=11)	FC08 (n=1)		Median: 161	Median: 82.5	750 (n=1)
		0.439 x 0.439 (n=1)	400 (n=1)	223 (n=1)	FC12 (n=2)		Mode: 150	Mode:75	

0.434 x							
0.434 (n=1)		230 (n=1)		FC43 (n=15)			
0.449 x							
0.449 (n=1)		233 (n=1)					
0.454 x							
0.454 (n=1)		238 (n=1)					
0.463 x							
0.463 (n=1)		239 (n=1)					
0.467 x							
0.467 (n=1)		240 (n=4)					

		0.468 x 0.468 (n=4)		246 (n=1)					
		0.481 x 0.481 (n=1)		250 (n=2)					
		0.488 x 0.488 (n=2)							
Philips (n=24)	2.00 (n=16)	0.404 x 0.404 (n=1)	221 (n=1)	207 (n=1)	B (n=23)	120 (n=24)	Range: 79 - 470	Range: 89 - 450	533 (n=1)
	2.5 (n=1)	0.432 x 0.432 (n=1)	222 (n=1)	221 (n=1)	C (n=1)		Median: 305.5	Median: 300	740 (n=1)

		0.684 x 0.684 (n=1)		350 (n=1)					
		0.977 x 0.977 (n=1)		500 (n=1)					
		1.168 x 1.168 (n=1)		598 (n=1)					
SIEMENS (n=29)	1.00 (n=4)	0.369 x 0.369 (n=1)	500 (n=23)	189 (n=1)	B30f (n=3)	120 (n=22)	Range: 75 - 560	Range: 77 - 400	500 (n=9)
	1.25 (n=4)	0.391 x 0.391 (n=1)	500.084 (n=5)	200 (n=1)	B30s (n=6)	130(n=5)	Median: 217	Median: 132	750 (n=9)

	3.00 (n=20)	0.402 x 0.402 (n=1)	Missing Data (n=1)	206 (n=1)	B31f (n=6)	140 (n=2)	Mode: 217	Mode: 122	800 (n=4)
	5.00 (n=1)	0.414 x 0.414 (n=1)		212 (n=1)	B31s (n=5)				1000 (n=6)
		0.418 x 0.418 (n=1)		214 (n=1)	B40s (n=4)				3362 (n=1)
		0.426 x 0.426 (n=1)		218 (n=1)	B70s (n=3)				
		0.434 x 0.434 (n=1)		222 (n=1)	H31s (n=1)				

0.451 x							
0.451 (n=1)		231 (n=1)	T20s (n=1)				
0.465 x							
0.465 (n=1)		238 (n=2)					
0.477 x							
0.477 (n=1)		244 (n=1)					
0.484 x							
0.484 (n=2)		248 (n=2)					
0.488 x							
0.488 (n=1)		250 (n=1)					

	0.498 x							
	0.498		255 (n=2)					
	(n=2)							
	0.500 x							
	0.500		256 (n=1)					
	(n=1)							
	0.508 x							
	0.508		260 (n=1)					
	(n=2)							
	0.518 x							
	0.518		265 (n=2)					
	(n=2)							
	0.533 x							
	0.533		273 (n=1)					
	(n=1)							

	0.561 x								
	0.561 (n=1)		287 (n=1)						
	0.563 x								
	0.563 (n=1)		288 (n=1)						
	0.572 x								
	0.572 (n=1)		293 (n=1)						
	0.586 x								
	0.586 (n=1)		300 (n=1)						
	0.596 x								
	0.596 (n=1)		305 (n=1)						

	0.604 x 0.604 (n=1)	309 (n=1)						
	1.000 x 1.000 (n=1)	512 (n=1)						
	1.367 x 1.367 (n=1)	700 (n=1)						

Table 4.1: Acquisition and reconstruction characteristics extracted from the DICOM header metadata of the CT images from Cohort 1. Recon = reconstruction

For the Cohort 1, with a fixed bin size of 10, 60% (n=26) of radiomic features showed poor reproducibility following spatial resampling with a %COV>50% and the maximum recorded %COV of 1884% for GLRLM_SRLGE. Fourteen percent (n=6) of radiomic features showed reasonable reproducibility with a %COV of <50%, but still >30% following resampling. The remaining 26% (n=11) demonstrated good reproducibility with a %COV<30%, with GLCM_Homogeneity being the most reproducible feature (%COV=1.7%) (Figure 4.2). Figure 4.3 shows these results ranked according to the %COV.

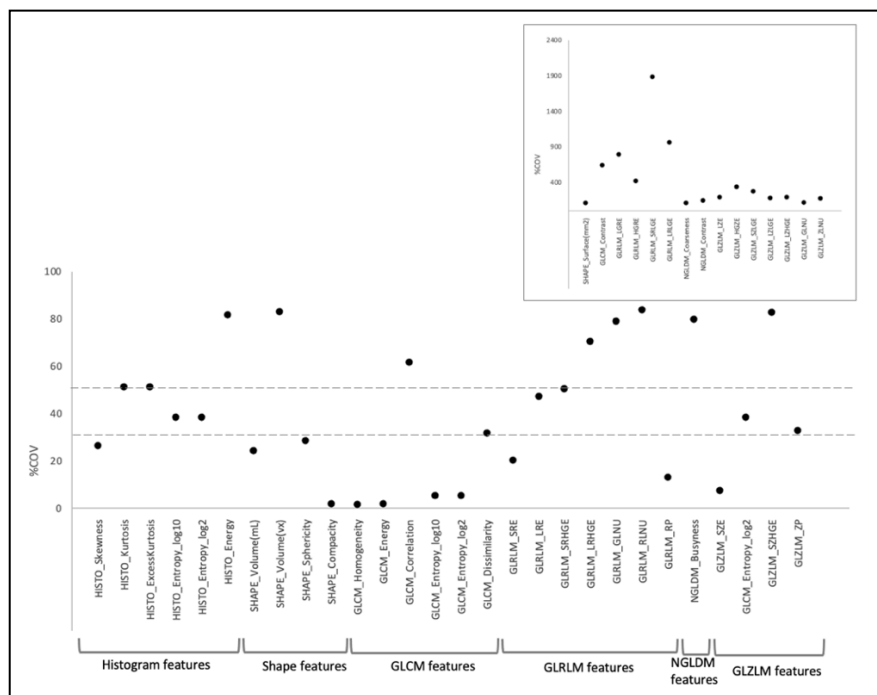


Figure 4.2: Variability of radiomic feature values following spatial resampling expressed as percentage coefficient of variance (%COV). Horizontal dotted lines indicate the 30% and 50% threshold lines. Radiomic features with %COV >100% are shown in the inset. **Data shown is based on a fixed bin size of 10.**

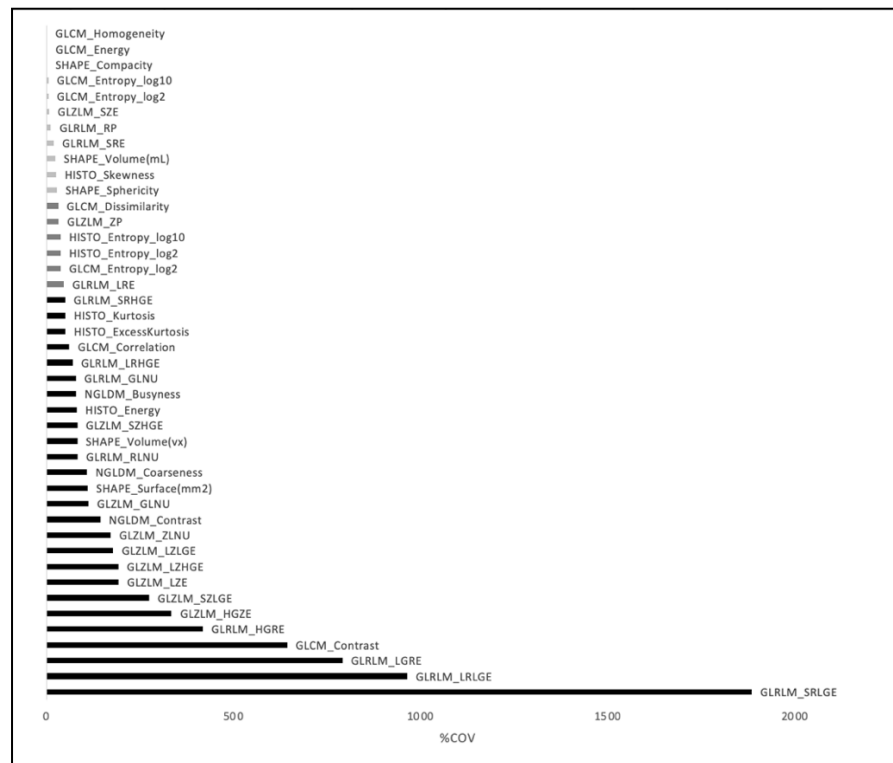


Figure 4.3: Variability of radiomic feature values following spatial resampling; ranked according to percentage coefficient of variance (%COV). Light grey %COV<30%, dark grey %COV <50% and >30%, black %COV >50%. Data shown is based on a fixed bin size of 10.

When the grey level discretization was adjusted to a fixed bin size of 5, overall improvement in reproducibility following spatial resampling was observed. Forty-two percent (n=18) of radiomic features had poor reproducibility, with the GLCM_Contrast showing the worst reproducibility (%COV=653%). Whilst GLRLM_SRLGE showed consistently poor reproducibility regardless of bin size used, the %COV for this feature dropped from 1884% to 64% for a bin size of 10 and 5 respectively; showing the impact discretisation can have on results. Again, 14% (n=6) of radiomic features showed reasonable reproducibility with a %COV between 50% and 30%; five of which had previously shown poor reproducibility when based on a bin size of 10 (HISTO_Kurtosis, HISTO_ExcessKurtosis, HISTO_Energy, SHAPE_Surface(mm2), GLZLM_SZLGE). The remaining 47% (n=20) radiomic features demonstrated good reproducibility following spatial resampling, based on a fixed bin size of 5 (Figure 4.4). Figure 4.5 shows the features ranked according to the %COV for a fixed bin size of 5.

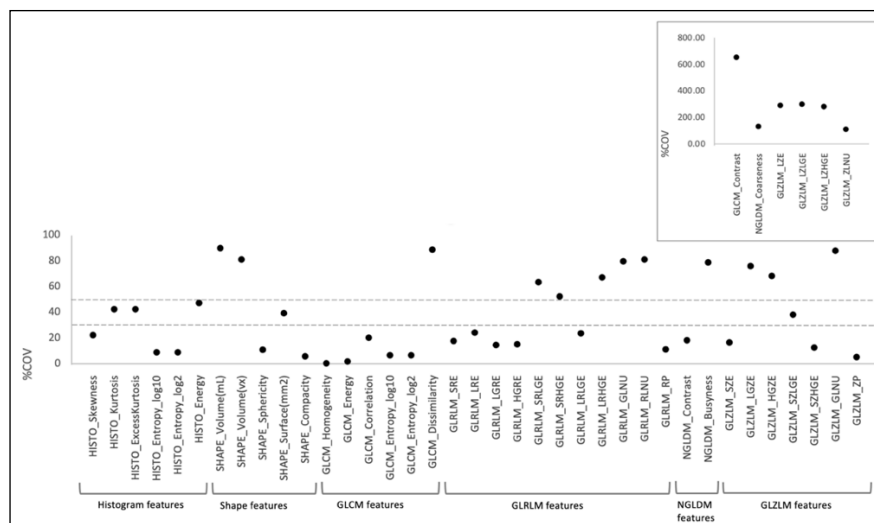


Figure 4.4: Variability of radiomic feature values following spatial resampling expressed as percentage coefficient of variance (%COV). Horizontal dotted lines indicate the 30% and 50% threshold lines. Radiomic features with %COV >100% are shown in the inset. **Data shown is based on a fixed bin size of 5.**

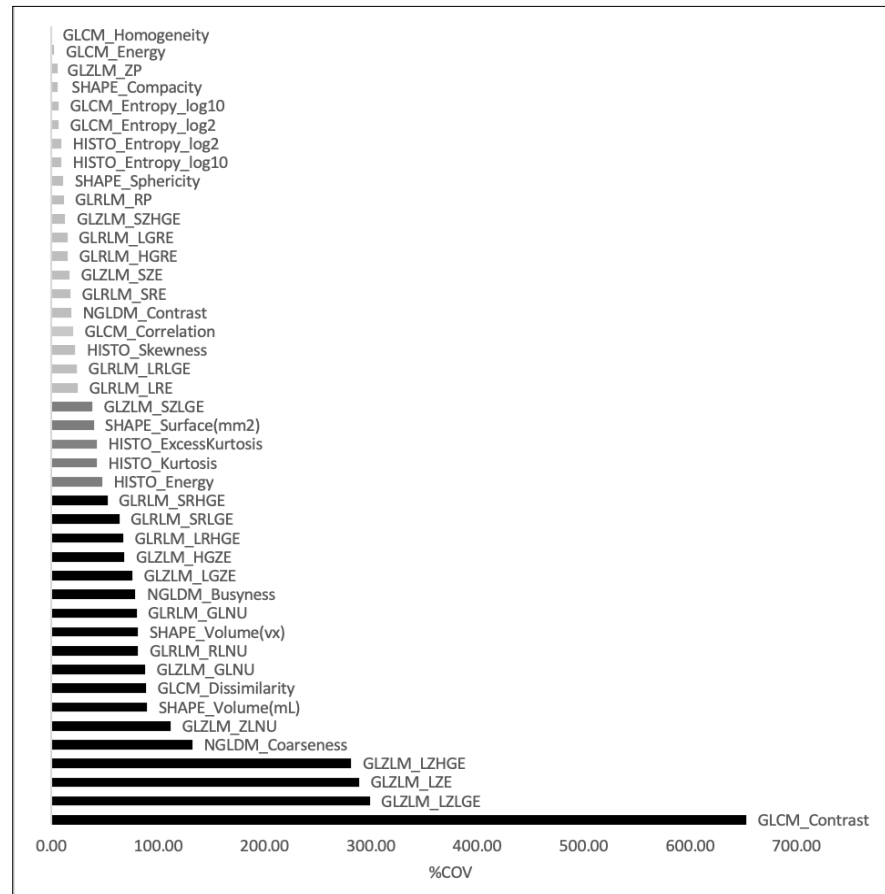


Figure 4.5: Variability of radiomic feature values following spatial resampling; ranked according to percentage coefficient of variance (%COV). Light grey %COV<30%, dark grey %COV <50% and >30%, black %COV >50%. Data shown is based on a fixed bin size of 5.

When comparing the impact of resampling for both bin sizes used, 35% of radiomic features experienced a reduction in the %COV that resulted in them being reclassified when a bin size of 5 was used. For example, for a bin size of 10, GLRLM_LGRE was considered to have poor reproducibility following resampling with a %COV of 792%; this value dropped to just 15% when the bin size was reduced to 5. Conversely, the %COV of 7% (n=3) of radiomic features increased when the bin size of 5 was used, resulting in reclassification; these were SHAPE_Volume(mL) with an increase of 66%, GLCM_Dissimilarity with an increase of 57% and, GLZLM_LGZE with an increase of 37% (Table 4.2, Figure 4.6).

Feature Family	Radiomic Feature	Fixed bin size of 10		Fixed bin size of 5		Consistency in classification	
		%COV	Class	%COV	Class		
HISTOGRAM	Skewness	26.57	1	22.30	1	Y	–
	Kurtosis	51.44	3	42.62	2	N	↓
	Excess Kurtosis	51.44	3	42.62	2	N	↓
	Entropy log10	38.45	2	9.19	1	N	↓
	Entropy log2	38.45	2	9.19	1	N	↓
	Energy	81.94	3	47.46	2	N	↓
SHAPE	Volume(mL)	24.32	1	89.94	3	N	↑
	Volume(vx)	83.09	3	81.30	3	Y	–
	Sphericity	28.75	1	11.18	1	Y	–
	Surface (mm2)	111.38	3	39.60	2	N	↓
	Compacity	2.08	1	6.07	1	Y	–
GLCM	Homogeneity	1.73	1	0.53	1	Y	–
	Energy	1.98	1	2.08	1	Y	–
	Contrast	643.65	3	652.82	3	Y	–
	Correlation	61.90	3	20.42	1	N	↓
	Entropy log10	5.43	1	6.71	1	Y	–
	Entropy log2	5.43	1	6.71	1	Y	–
	Dissimilarity	31.93	2	88.83	3	N	↑
GLRLM	SRE	20.33	1	17.75	1	Y	–
	LRE	47.36	2	24.41	1	N	↓
	LGRE	791.68	3	14.86	1	N	↓
	HGRE	418.55	3	15.29	1	N	↓
	SRLGE	1884.90	3	63.60	3	Y	–
	SRHGE	50.56	3	52.43	3	Y	–
	LRLGE	964.48	3	23.87	1	N	↓
	LRHGE	70.58	3	67.37	3	Y	–
	GLNU	79.12	3	79.88	3	Y	–

	RLNU	83.84	3	81.37	3	Y	–
	RP	13.10	1	11.35	1	Y	–
NGLDM	Coarseness	108.84	3	132.33	3	Y	–
	Contrast	144.86	3	18.48	1	N	↓
	Busyness	79.85	3	78.87	3	Y	–
GLZLM	SZE	7.60	1	16.70	1	Y	–
	LZE	193.16	3	289.22	3	Y	–
	LGZE	38.64	2	76.07	3	N	↑
	HGZE	335.29	3	68.44	3	Y	–
	SZLGE	274.75	3	38.37	2	N	↓
	SZHGE	82.88	3	12.87	1	N	↓
	LZLGE	179.26	3	299.12	3	Y	–
	LZHGE	192.31	3	281.62	3	Y	–
	GLNU	113.33	3	87.97	3	Y	–
	ZLNU	172.43	3	111.75	3	Y	–
	ZP	33.01	2	5.46	1	N	↓

Table 4.2: Percentage coefficient of variance (%COV), along with classification of reproducibility of radiomic features following spatial resampling of Cohort 1 (n=127). Class 1 %COV <30%; Class 2 %COV >30% but <50%; Class 3 %COV >50%. Consistency in class across both bin sizes investigated is also noted.

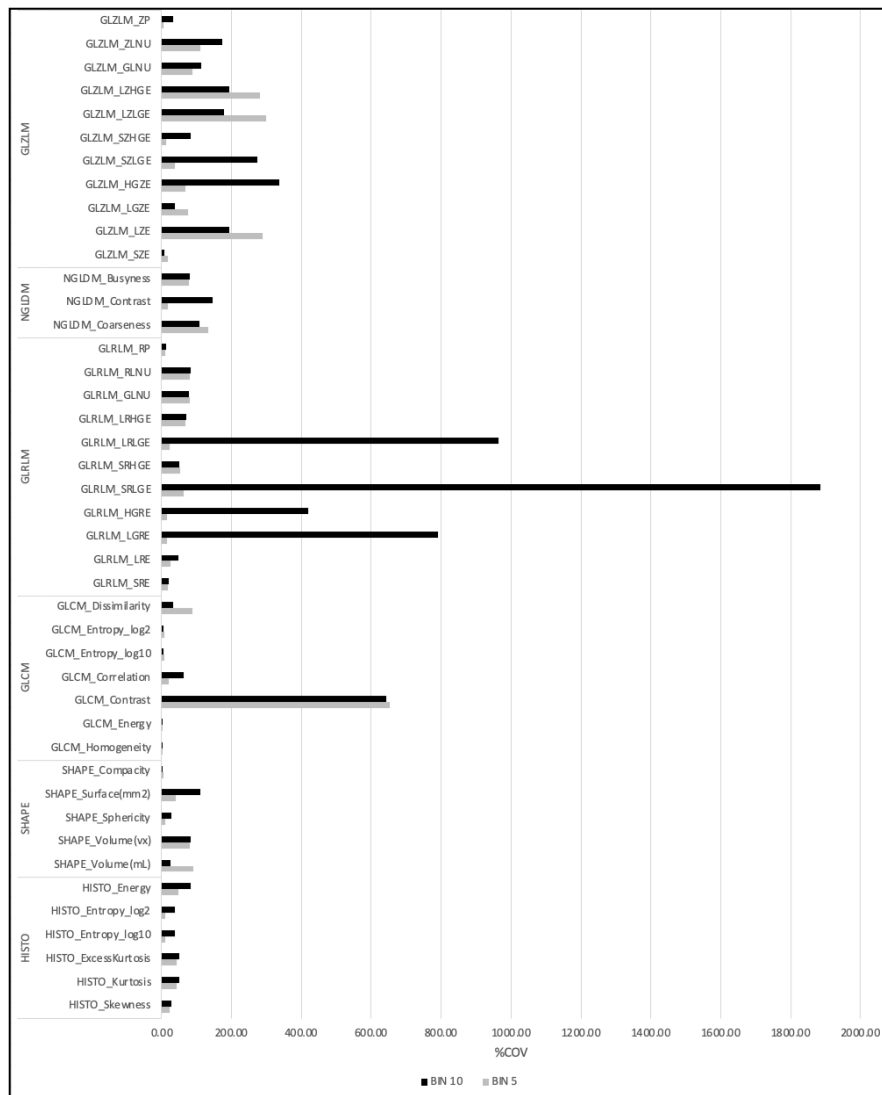


Figure 4.6: Variability of radiomic feature values following spatial resampling expressed as percentage coefficient of variance (%COV). Data shown is based on a fixed bin size of 5 (grey) and 10 (black).

4.3.2 Cohort 2

The impact of spatial resampling based on a fixed bin size of 5 was also confirmed based on the results of the second cohort (Table 4.3). Fifty-six percent (n=24) of radiomic features were considered to be reproducible (%COV<30%) and 33% (n=14) showed poor reproducibility (%COV>50%). Fifty-three percent (n=23) of radiomic features remained in the same reproducibility class for both Cohort 1 and 2.

Feature Family	Radiomic Feature	Cohort 1		Cohort 2		Consistency in Classification	
		%COV	Class	%COV	Class		
HISTOGRAM	Skewness	22.30	1	33.90	1	Y	–
	Kurtosis	42.62	2	52.70	3	N	↑
	Excess Kurtosis	42.62	2	52.70	3	N	↑
	Entropy_log10	9.19	1	70.32	3	N	↑
	Entropy log2	9.19	1	7.97	1	Y	–
	Energy	47.46	2	70.76	3	N	↑
SHAPE	Volume(mL)	89.94	3	34.18	2	N	↓
	Volume(vx)	81.30	3	24.67	1	N	↓
	Sphericity	11.18	1	7.86	1	Y	–
	Surface (mm2)	39.60	2	28.09	1	N	↓
	Compacity	6.07	1	8.70	1	Y	–
GLCM	Homogeneity	0.53	1	8.52	1	Y	–
	Energy	2.08	1	68.21	3	N	↑
	Contrast	652.82	3	27.69	1	N	↓
	Correlation	20.42	1	6.17	1	Y	–
	Entropy log10	6.71	1	12.62	1	Y	–
	Entropy log2	6.71	1	12.62	1	Y	–
	Dissimilarity	88.83	3	7.34	1	N	↓
GLRLM	SRE	17.75	1	22.31	1	Y	–
	LRE	24.41	1	34.19	2	N	↑
	LGRE	14.86	1	15.27	1	Y	–
	HGRE	15.29	1	15.59	1	Y	–
	SRLGE	63.60	3	11.99	1	N	↓
	SRHGE	52.43	3	10.85	1	N	↓
	LRLGE	23.87	1	12.36	1	Y	–
	LRHGE	67.37	3	36.75	2	N	↓
	GLNU	79.88	3	22.13	1	N	↓
	RLNU	81.37	3	31.01	2	N	↓
	RP	11.35	1	20.59	1	Y	–

NGLDM	Coarseness	132.33	3	22.26	1	N	↓
	Contrast	18.48	1	19.82	1	Y	–
	Busyness	78.87	3	114.80	3	Y	–
GLZLM	SZE	16.70	1	24.35	1	Y	–
	LZE	289.22	3	96.48	3	Y	–
	LGZE	76.07	3	126.56	3	Y	–
	HGZE	68.44	3	76.95	3	Y	–
	SZLGE	38.37	2	996.84	3	N	↑
	SZHGE	12.87	1	57.92	3	N	↑
	LZLGE	299.12	3	86.88	3	Y	–
	LZHGE	281.62	3	107.42	3	Y	–
	GLNU	87.97	3	44.05	2	N	↓
	ZLNU	111.75	3	67.18	3	Y	–
	ZP	5.46	1	5.91	1	Y	–

Table 4.3: Percentage coefficient of variance (%COV), along with classification of reproducibility of radiomic features following spatial resampling of Cohort 1 (n=127) and Cohort2 (n=17). Class 1 %COV <30%; Class 2 %COV >30% but <50%; Class 3 %COV >50%. Consistency in class between Cohorts 1 and 2 is also noted. Note: results are based on fixed bin size of 5 for both groups.

In summary, following spatial resampling, 58% (n=25) of radiomic features demonstrated consistent results regardless of either bin size used. Of these, the classification of 67% (n=16) of the features were confirmed based on Cohort 2. Six features showed consistent poor reproducibility with a %COV>50% (NGLDM_Busyness, GLZLM_LZE, GLZLM_HGZE, GLZLM_LZLGE, GLZLM_LZHGE, GLZLM_ZLNU), and 10 showed consistently good reproducibility with a %COV<30% (HISTO_Skewness, SHAPE_Sphericity, SHAPE_Compacity, GLCM_Homogeneity, GLCM_Entropy_log10, GLCM_Entropy_log2, GLRLM_SRE, GLRLM_RP, GLZLM_SZE, GLZLM_ZP).

4.4 DISCUSSION

The primary aim of this study was to quantify the impact of spatial resampling of CT images on radiomic features using the whole left parotid gland as the ROI. The reproducibility of these results were then tested using two different fixed bin sizes to observe the impact of grey level discretization on results. Unique to this work is the lack of phantom component, where the analysis was performed using a patient dataset from TCIA. This was purposefully done to reflect the challenges when working with real patient data, in which the texture composition of soft tissue can be remarkably different to that of a phantom. This approach is aligned with the recent call made by Traverso et al., (137) to conduct benchmarking studies based on “common, shared, and publicly available datasets”, such as that used in this study. Furthermore, by using images taken from a heterogeneous cohort, the results presented here better reflect what is often seen when working with data collected across multiple departments where variation in image acquisition and reconstruction protocols is common. Finally, this work also included two separate cohorts to evaluate the robustness of conclusions drawn.

4.4.1 Impact of spatial resampling

CT scan parameters, such as tube voltage, pixel size and slice thickness, are often adjusted on a case-by-case basis in order to balance image quality and dose received by the patient. Whilst clinically justified, this leads to considerable variability in CT parameters when working with large patient cohorts, as seen in Cohort 1. It is understood that this variability will impact on radiomic texture feature values, given radiomic features convey spatial and intensity information of an image (137, 180, 187), and resampling to a common isotropic voxel grid may reduce inter-acquisition variations in radiomic feature values caused by differences in both in-plane and through-plane spatial resolution (188, 189). A number of studies have investigated

this impact of spatial resampling based on MR, PET, and CT images using either phantoms, real patient data, or a combination of both (137).

Existing phantom studies which specifically assessed reproducibility of features based on CT imaging, all concluded that feature reproducibility is strongly affected by voxel size resampling (137). One of the earliest phantom studies in this field is the work of Shafiq-ul-Hassan et al. (179), where it was determined that 30% of features were highly sensitive to the voxel size. Following resampling to an arbitrarily chosen common voxel size ($1 \times 1 \times 2 \text{ mm}^3$), they were able to demonstrate improved robustness of 20% of features; leading them to recommend resampling as a way to eliminate dependencies introduced by voxel size variability. Later the same year, Mackin et al., (187) published a follow up study to their earlier work (180) which included a small patient validation cohort, using the lung tumour as the region of interest. This work also now included the combination of resampling along with image filtering to further reduce variability of feature values. Whilst resampling was successful for the phantom component of the study, resampling alone actually increased variability in the patient cohort. Whilst this result was unexpected, it is not particularly informative given the unacceptably small patient numbers in their study cohort ($n = 8$).

Where phantom studies have the luxury of rescanning the phantom multiple times with varying slice thickness and reconstruction sizes, prospective patient studies cannot ethically do this due to unjustified increased dose to the patient. Furthermore, most radiomic studies using real patients have been retrospective in nature; therefore relying on either having access to the raw CT to allow for varying reconstruction parameters to be investigated, or (in the case of using data mined from imaging repositories) resampling the in-plane spacing and slice thickness, to investigate the influence of different pixel and voxel sizes. The former point (i.e. having access to preserved/unreconstructed CT data) was the contributing factor to the low patient numbers ($n = 8$) cited in the seminal work by Mackin et al., (187) previously discussed. Regardless, there are relevant studies based on patient data and although sample sizes have

increased, the majority of them are still <100 patients with some including a mix of cancer sites within the one cohort (137). One such study is that by Bogowitz et al., (190) in which their patient cohort consisted of patients with either lung cancer (n=11) or head and neck cancer (n=11). Based on radiomic analysis of CT perfusion maps, they investigated feature robustness under a number of image processing conditions (including voxel resampling). Of all the potentially standardizable imaging parameters, the highest feature variability was observed for voxel size, in which an incredible 98% of features in the head and neck cohort were considered unstable.

It is widely accepted that voxel size is hugely influential in radiomic feature values for CT imaging (137, 190). What is less clear is the determination of an appropriate voxel size for the clinical scenario at hand. In their phantom study, Larue et al., (181) resampled the images to a voxel size of 1x1x3mm, as they felt this corresponded to the voxel sizes used in most clinical images. Shafiq-Ul-Hassan et al., (179) arbitrarily chose a common voxel size of 1x1x2mm for the purpose of spatial resampling. Isotropicity, however, is important for texture analysis (125); and resampling to a common isotropic voxel of 1x1x1mm as in this study, while indiscriminate, is not uncommon (185, 187, 191). When attempting to harmonise the voxel size within a mixed cohort, Ibrahim et al., (188) suggest that the mode voxel value should be used, resulting in some patient images being up-sampled and others being down-sampled following spatial resampling. With respect to in-plane spacing, interpolating from smaller pixels to larger pixels occurred in the majority of cases in Cohort 1, where the mode pixel size was originally 0.488mmx0.488mm. Degrading the spatial resolution may actually hold a number of advantages, for example reducing noise. This increase in spatial resolution also means the resampling process is less sensitive to the method of interpolation used where “new” pixels do not need to be created, as is the case when images are up-sampled (188).

With respect to interpolation methodologies used for resampling, 3D interpolation is the most common approach due to its relative simplicity;

however, this study made use of the non-linear (polynomial lagrangian interpolation) method available within LIFEx (123). Irrespective of the idiosyncrasies of interpolation methods (186), or the voxel size chosen, resampling to a common voxel size is a necessary step in order to bring all the images into the same frame of reference (181). Furthermore, Bogowicz et al., (190) highlight the importance of isotropic voxel sizing considering many radiomic feature definitions assume symmetry in voxel size; a sentiment which is also outlined in the Imaging Biomarker Standardisation Initiative (125). Interestingly of the radiomic studies that analysed the parotid gland as the region of interest, Gabrys et al., (185) was the only one to resample the CT images to a common isotropic voxel size, with other studies accepting anisotropic voxel sizes (116, 177, 184).

4.4.2 Impact of Grey Level Discretization

Discretization is the process in which original grey level values are grouped based on a specific range intervals (i.e. bins) in order to reduce in particular the impact of noise on the radiomic features (135). Whilst there is an understanding that grey-level discretization has an impact on radiomic feature values, there is a lack of consensus in the literature regarding the most appropriate bin width to be used, particularly in head and neck studies (116). Van Dijk et al., was one of the earliest studies to apply radiomic analysis with the parotid gland as the region of interest. Using in-house developed software, they extracted radiomic features based on a bin size of 25; although no rationale was provided for this choice in bin size. Sheikh et al., (116) also investigated the parotid gland as the region of interest and also used a bin width of 25, citing previous head and neck publications as their guide (113, 192). Whilst being guided by previous literature is a reasonable approach, inherent differences in methodology with the studies they cited must be noted; for example, the imaging modality (MRI vs CT), the ROI considered (tumour vs normal tissue), and even the radiomics software used (PyRadiomics in Python vs LIFEx).

As mentioned, Sheikh et al., (116) cited Thor et al., (192) as supporting literature for their discretization methodology. Thor et al., (192) analysed MR images, and as such used a fixed number of bins rather than a fixed bin width common to CT based radiomic studies. This approach is supported by the IBSI for scenarios (such as MRI) in which the intensity units are arbitrary (125). Whilst Sheikh et al., also used MRI (combined with CT) in their study, a fundamental difference that must be acknowledged is the difference in the ROI examined. Thor et al., (192) assessed trismus as the clinical endpoint, and therefore analysed radiomic features within the masseter, lateral and medial pterygoids and temporalis muscles, whereas Sheikh et al., (116) built a predictive model for xerostomia and as such the ROIs in their study were the parotid and submandibular glands. In addition, Sheikh et al., also cited Leijenaar et al., (113) to support their discretization methodology. Using in-house developed software, Leijenaar et al.,(113) developed a radiomic signature to predict HPV-16 status based on analysis of CT imaging, and in doing so considered that the use of different bin sizes may provide additional information. For that reason, they conducted their analysis based on a bin width of both 25 and 10. Following feature reduction based on pairwise correlations, no preference in bin size was observed. Both Sheikh et al., (116) and Leijenaar et al., (113) used in-house developed software based on python and Matlab respectively, but again the ROI differed in each study (i.e. gross tumour volume versus parotid/sub mandibular glands).

Similar to this work, Nardone et al., (129) used LIFEx radiomic software to assess textural features of the parotid gland based on CT imaging; however, they did not publish specific details, such as discretization approaches in their methodology. Sharing such details would be welcomed, thereby enabling easier comparison of outcomes across studies, which is a current limitation in many head and neck radiomic studies (193-195).

This study assessed the reproducibility of radiomic features based on two different bin widths. For the majority of radiomic features (58%), the classification of reproducibility was consistent regardless of whether a bin

width of five or ten was used; implying that these features are not influenced by grey-level discretization. That being said, the variability of 35% of features in this study saw an improvement in their classification based on the smaller bin width. Just 11 features were considered highly reproducible (i.e. Class 1, %COV<30) with a bin width of ten, yet this figure rose to 20 when the bin width was reduced to 5. Interestingly, under the more controlled environment of a phantom study, Shafiq-ul-Hassan et al., (179) found just seven out of 51 texture features were reproducible (%COV<20) with varying grey level discretization. When compared to results presented here, just GLRLM_RP and GLRLM_SRE also had %COV≤20% irrespective of bin width. Remarkably NGLDM_Coarseness showed particularly poor reproducibility in this study in which the %COV was in excess of 100% for both bin widths, compared to the excellent (%COV<20%) independent reproducibility noted by Shafiq-ul-Hassan et al (179). Possible explanations for these contrasting results could be the fact that their imaging was based on a phantom, calling on a need for real patient studies to allow for more meaningful comparisons. Shafiq-ul-Hassan et al., (196) did also conduct a follow up study in which 18 patients with non-small lung cancer were examined. Based on the analysis of 17 specific texture features, they were able to validate the results from their previous phantom study. Although the use of real patient data is a welcome addition to the literature, results cannot be extrapolated to different clinical sites. For example, Shafiq-ul-Hassan et al., (196) reported the GLRLM_GLNU texture feature to be robust irrespective of bin sizes used (ICC=0.9); however, results presented in Study 2 conflict with this, in which GLRLM_GLNU was not found not to be reproducible (%COV=79% for both bin widths). Fundamental differences in tissue composition of lung tumours compared to healthy parotid glands help to explain these results. Lung tumours are known to be heterogeneous in composition caused by differences in genetic mutations, metabolic, histological and cellular heterogeneity, and, variability in the presence of hypoxia (197, 198). Conversely, although the parotid gland contains two separate lobes, the parotid duct, and the facial nerve, it is largely composed of adipose tissue

with serous secreting cells; all of which are relatively homogenous. Considering the GLRLM gives the size of the homogenous runs for each grey level, and the GLRLM_GLNU depicts the “non-uniformity” of these grey level runs, it is not surprising to see these differences in results for lung tumours compared to healthy parotid glands. These differences in results can be further rationalised based on inherent differences in radiomics software, where the precise calculation of textural features can vary (123).

Using the same CCR phantom as Shafiq-ul-Hassan et al. (179), Larue et al., (181) also found that bin width can potentially have a large impact on feature values; an observation which was less pronounced when using a bin width of 25, compared to using a smaller bin width of 5 or a larger bin width of 50. Despite these observations, they concluded that discretization did not impact on the stability of features overall; a result which was supported by our data where consistency in classification was observed in 58% of radiomic features.

Considering that defining the bin width is a necessary step in building a histogram (123, 133), features derived from the HISTO family may be dependent on the bin width selected. In this study the %COV for the HISTO_Energy feature was nearly halved when the smaller bin width was applied; with the %COV of 82% and 48% for bin widths of 10 and 5 respectively. Knowing that HISTO_Energy reflects the uniformity of the distribution and HISTO_Entropy reflects the randomness of the distribution, it is not surprising to see these features are similarly influenced by discretisation. This was observed in the results where the %COV for HISTO_Entropy fell from 38% for the bin width of 10 to 9% for the bin width of 5. Conversely, the %COV for the HISTO_Skewness feature had just a 4% difference between the two different bin widths. This leads us to conclude that HISTO family features are influenced by bin width to different extents depending on the specific feature analysed.

Another feature family with a single outlier in the classification results between different bin sizes was the NGLDM family. Here both the NGLDM_Coarseness and NGLDM_Busyness were not influenced by bin size

with a consistent %COV>50%; yet the NGLDM_Constrast %COV fell from 145% to 19% when the bin size was adjusted from 10 to 5. This may also be rationalised by looking at the classic definition of these specific features, where the NGLDM_Coarseness and NGLDM_Busyness refer to the “spatial rate of change” and “frequency of change” respectively; whereas the NGLDM_Constrast simply refers to the “intensity difference between neighbouring regions”(133). The remaining feature families all showed inconsistent results within them, where the classification of some specific features was reproducible across the different bin widths, and others weren't. When conducting texture analysis using radiomics this demonstrates the need to consider each radiomic feature independently, and whilst feature reduction may be a necessary step for some radiomic software, the low number of features extracted from LIFEx makes this level of analysis manageable. However, some may argue when working with fewer textural features there is a risk that some potentially informative data are missing.

Although the results presented in this study indicate the smaller bin width of 5 will result in less variability in radiomic feature values following spatial resampling, this does not infer that further reducing the bin width will improve these results even more. In fact if the bin width is too small, the discretization step is essentially made redundant due to the difficulties in distinguishing what is genuine textural difference of the underlying tissue and what is simply noise from the image itself (135). Whilst the question regarding the optimal bin width remains unanswered, van Timmeren et al., (135) suggests first inspecting the histogram distribution of HU for ROI to be used, and then selecting an appropriate bin width based on this information. In this study, initially the default bin width of 10 from the LIFEx software was used. This was then halved to 5 in case texture analysis was missed due to too few bins being used. Although there are no clear recommendations regarding the optimal bin width, we can be guided by the variation in HU that is often observed and accepted in routing quality assurance. For example, for a CT acquired at 120kV, the HU in a water

phantom are 0 +/- 6HU, so binning in values less than 5 HU may not be appropriate.

The majority of other patient studies in which feature reproducibility was assessed were either based on lung tumours, non-CT imaging, or the image pre-processing investigated focused on filtering rather than spatial resampling (137), limiting direct comparison with results presented here.

In addition to the recognized limitations regarding discretization, there are other methodological limitations of this study which must be noted. Defining the voxel intensity range (or re-segmentation range as it is referred to in the IBSI) is a necessary step prior to feature extraction (199). This is the process in which the range of HUs used in the texture analysis are defined, and pixels/voxels which fall beyond a defined range of HUs are excluded from the ROI for analysis (135). As per the IBSI recommendations for CT imaging, this is typically based on a “half-open interval (a,∞) ” or “closed interval (a,b) ” (125). In Study 2, the latter approach was adopted in which the range of grey levels (i.e. HUs) was limited to reflect the minimum and maximum HU of the left parotid gland across the whole cohort; that is, a range of -1000 – 1800 for Cohort 1, and -561 – 1632 for Cohort 2. These values reveal that there were evidently voxels within the ROI that included areas of high and low dense material not reflective of parotid glandular tissue, most likely bone, and air. Whilst this is not ideal, the workflow in this study does reflect common clinical practice and it is possible that this could be observed in other scenarios. In future radiomic studies, these errors could be corrected for by using intensity outlier filtering. Regardless, choices regarding intensity filtering may potentially influence results, and to date has remained largely under-reported in radiomic studies (199).

Finally, this work consisted of just two cohorts; the purpose of the second one was simply to confirm the results observed. It could be argued that a third cohort would be useful to test generalisability of results.

4.4.3 Clinical Application

This study has added to the literature on the impact of spatial resampling and grey level discretisation on reproducibility of radiomic features, which is largely limited to either phantom studies or lung tumour volumes. To date, no other study has examined these influences using the whole parotid gland as the region of interest. What is lacking in this study current work is a clinical application of these results (i.e. integrating radiomics in the predictive models for xerostomia), and based on this study, it is not known how these observations would impact on the performance of a regression or classification model. Where there is no prognostic value component considered, it is not possible to clearly state that one bin width is superior for this specific scenario. On the other hand, studies that do contain prognostic modelling based on radiomics, may be basing results on features that are either not reproducible or on a model trained on a highly homogeneous dataset. Whilst the discriminative power and reproducibility of radiomic features cannot be considered in isolation, results presented in this study adds to the current understanding of radiomic features for the prediction of xerostomia (200).

In 2017, van Dijk et al., (184) clearly isolated the GLRLM_SRE feature as being associated with the development of moderate to severe xerostomia measured at 12 months (XER_{12m}) following treatment completion. This result was based on the improved overall and discrimination power of the model when this feature was included (Likelihood Ratio Test: $p=0.01$). The data from Study 2 adds weight to this, considering this feature was shown to be reproducible following spatial resampling irrespective of bin width used. This result was also confirmed based on the second cohort. Also within the GLRLM feature family, based on univariable analysis, van Dijk et al., (184) noted that GLRLM_LRE and GLRLM_SRHGE were significantly correlated to XER_{12m} ; however, these features were not later confirmed in the multivariable analysis, as they were not isolated by the lasso modelling process. Sheikh et al., (116) also incorporated radiomics into the development of a number of xerostomia predictive models. Their study also included the submandibular glands, as well as MRI in addition to CT

imaging. For the model based on CT imaging only, they were able to confirm previous work in this field in which GLRLM features were isolated as predictive of xerostomia. They also highlighted GLSZM features to enhance the performance of the model. From this work it is noted that most features derived from the equivalent GLZLM family were consistently non-reproducible, and as such should be applied with caution without further harmonization. Notably, their work also explored more comprehensive models. These models which were generated based on the combination of MRI along with the standard CT, proved to have improved predictive power. One limitation of their work, which is relevant here, is the lack of resampling of the CT images and the use of non-cubic voxels. The authors acknowledge that this is an area of active research and that resampling to an isotropic voxel size may improve interpretation of radiomic features.

4.5 CONCLUSION

The application of radiomics for the prediction of xerostomia is an active area of research. The work presented here was prompted by a need to examine radiomic feature reproducibility following spatial resampling and grey level discretisation. These two necessary components have been overlooked in radiomics research where the parotid gland is studied. This work has identified a number of radiomic features which are reproducible to such conditions and will strengthen the current body of evidence on this topic.

5.1 INTRODUCTION

The EORTC-REACH phase III randomised control trial investigated the superiority of avelumab in combination with cetuximab and radiotherapy over the standard of care which uses cisplatin-RT or cetuximab-RT alone for the treatment of stage III/IV squamous cell carcinoma of a primary located in oral cavity, oropharynx, hypopharynx or larynx (32). Based on a safety analysis of 41 patients randomised to either of the experimental arms, 61% experienced some degree of acute xerostomia, and 22% of patients experienced grade II or grade III acute xerostomia, measured within the first eight weeks of the commencement of treatment (30). These outcomes are consistent with previous studies, reiterating the incidence and impact of radiation induced xerostomia, even in the era of image guided intensity modulated radiotherapy (201). Although the incidence of sticky saliva was not reported on in the preliminarily safety analysis by Tao et al., (30) given the link with xerostomia, it is likely that similar rates would have been reported.

Traditional normal tissue complication probability models, which relied on dose-volume histogram metrics that lack spatial information, predicted the risk of late rather than acute xerostomia (47). Although a dose-response relationship of the whole parotid glands was established, the prediction of xerostomia may also be limited by the inability of these traditional models to account for the complexity of the gland's architecture and its response to radiation (202). Furthermore, many of these models were derived from quantitative assessment of salivary flow, thereby failing to consider the subjective nature of xerostomia and its impact on the patient's quality of life. Quantitative imaging analysis, through radiomic feature extraction

from CT images, can reveal the complexities of parotid gland architecture; for example, variation in image intensity or subtle textural patterns that are not appreciated with the naked eye. Uncovering these previously “unseen” imaging biomarkers has the potential to improve the accuracy of predictive models (203).

The central hypothesis of this chapter is that radiation-induced xerostomia can be predicted prior to the manifestation of clinical signs, using radiomic analysis of routine CT images acquired for radiotherapy planning purposes. This study aimed to improve the prediction of acute xerostomia and sticky saliva (SS) through the inclusion of radiomic analysis of the parotid glands for patients with head and neck cancer.

5.2 MATERIALS AND METHODS

5.2.1 Patient Cohort and Treatment Protocol

Following ethical approval from the School of Medicine Research Ethics Committee (SOMREC), Trinity College Dublin (No: 20210903), anonymous data from 98 patients previously treated for head and neck cancer on the EORTC REACH phase III randomised control trial at Gustave Roussy were made available (30). Local ethics for Gustave Roussy was handled by the research collaborators in France, as stipulated in both the Data Protection Impact Assessment (DPIA) at TCD as well as the SOMREC ethics application.

Patients were treated five days per week using an IMRT simultaneous integrated boost technique to a prescribed dose of 69.96 Gy in 33 fractions (2.12 Gy/fraction) to the high-risk disease and 52.8 Gy (1.6 Gy/fraction) prophylactically to the low-risk region.

For the purpose of this study, anonymous data collected included patient age, sex, diagnosis, disease stage, HPV status (where appropriate), history of smoking and alcohol consumption, chemotherapy agents, radiotherapy

total dose, and mean dose to the left and right parotid glands. Patient age and mean parotid doses were left as continuous data, whilst sex, history of smoking, alcohol consumption, and HPV status were coded to binary values.

5.2.2 Quality of Life Assessment

As per the trial protocol (30), quality of life data were collected using the EORTC QLQ-HN35 (143) at baseline (i.e. at the time of treatment consent), 3 months, 15 months, and 27 months following treatment. Every patient received 33 fractions, treated 5 days per week with total fractionation time of 6 and a half weeks. Using this information, in addition to baseline, the EORTC QLQ-HN35 questionnaire (143) should have been completed approximately 5, 17, and 29 months after baseline.

For the purpose of this study, data collected within 5 months of baseline was defined as acute toxicity, data between 5- and 17-months following baseline was defined as sub-acute toxicity, and data collected greater than 17 months following baseline was defined as late toxicity.

The outcomes considered in this study were xerostomia and sticky saliva, and as such responses to question 42 (“Have you had a dry mouth”) and question 43 (“Have you had sticky saliva”) from the EORTC QLQ-HN35 questionnaire were collected (143). The event of xerostomia was dichotomised as no/mild xerostomia reported (Q41 score = 1-2) and moderate/severe xerostomia reported (Q41 score = 3-4). The presence of sticky saliva was also dichotomised in the same fashion.

If a patient had more than one questionnaire completed during a given time period, the highest score was taken, to capture the most severe level of toxicity experienced by the patient during that period.

5.2.3 CT Images and Radiomics Feature Extraction

Metadata from the planning CT scans of the 98 patients were extracted using OsiriX Lite (version 13.0.2). The scans were all acquired on a Siemens CT scanner with some variation in acquisition and reconstruction parameters (Table 5.1).

Manufacturer	SIEMENS
Manufacturer Model Name	Sensation Open or SOMATOM go.Sim
Slice Thickness	2mm
Tube Voltage (kVp)	120
Data Collection Diameter	500mm or 600.5mm
Software Versions	Syngo CT 2014A Syngo CT VA30A
Reconstruction Diameter (FOV)	Range: 482 – 820mm
Exposure Time	1000 or 1250
Xray Tube Current (mA)	Range: 54 – 399
Exposure	Range: 51 – 532
Focal Spots	1.2
Convolution Kernel	B31s or Br36f
Pixel Spacing	Range: 0.94140625 x 0.94140625 - 1.6015625 x 1.6015625
Rows	512
Columns	512

Table 5.1: Summary of Metadata from the planning CT scans extracted using OsiriX Lite (version 13.0.2).

Radiomic analysis was conducted using LIFEx software, version 7.3, and HU were binned using a fixed interval of 5 HU from -1024 to 1739, resulting in 553 bins. The choice of adopting a bin width of 5 HU was based on the findings presented in Study 2, in which the smaller bin width of 5 HU resulted in less variability in radiomic feature values following spatial resampling. As in Study 2, the values used for intensity range were based on the minimum and maximum range present in the patients' parotid glands. Prior to radiomic feature analysis, all images were resampled to a common voxel size of 1x1x1mm. The feature extraction process was also repeated based on a bin size of 10, totalling 227 bins, to allow a comparison of results across different bin sizes.

The feature extraction was run to report on 44 first order features (14 shape and 30 intensity histogram) and 56 texture features.

5.2.4 Testing Similarity in Radiomic Features Between Ipsilateral and Contralateral Parotid Glands

To determine the difference between the radiomic features extracted from patients' ipsilateral and contralateral parotid glands, a Wilcoxon signed rank test was performed. This process involved first coding each patient's left or right parotid gland as either ipsi- or contralateral based on location of the primary gross tumour volume (GTV_T) relative to midline. Radiomic feature values for the ROIs per patient were then separated and standardized using z-score normalisation using the following formula: $(x - \mu)/\sigma$, in which x is the original feature value, μ is the mean value of the feature based on the cohort, and σ is the standard deviation of the original feature values. Normalisation of features is required to align radiomic features into the same frame of reference and serves to standardise features which may vary due to differences in imaging acquisition and reconstruction protocols.

The results of this interim analysis confirmed that the radiomic feature values for the ipsilateral and contralateral parotid glands were similar per

patient ($p>0.05$; Table 5.2 and 5.3). Consequently, the pairs of original, un-normalised radiomic features extracted from each parotid gland per patient were averaged, resulting in a single value per feature for each patient. This approach was deemed appropriate considering no patient had a primary parotid tumour. These averaged features were then normalised using z-score normalisation.

Feature Family	Radiomic Feature	Z	p value	Effect Size
SHAPE	Volume	-0.470 ^b	0.638	-0.057
	Approximate Volume	-0.477 ^b	0.634	-0.058
	Voxels Counting	-0.477 ^b	0.634	-0.058
	Surface Area	-0.171 ^b	0.864	-0.021
	Surface To Volume Ratio	-0.006 ^c	0.995	-0.001
	Compacity	-0.183 ^b	0.855	-0.022
	Compactness1	-0.018 ^c	0.985	-0.002
	Compactness2	-0.031 ^c	0.976	-0.004
	Spherical Disproportion	-0.147 ^b	0.883	-0.018
	Sphericity	-0.031 ^c	0.976	-0.004
	Asphericity	-0.147 ^b	0.883	-0.018
	Centre Of Mass Shift	-4.467 ^c	0.000	-0.542
	Maximum3DDiameter	-0.092 ^c	0.927	-0.011
	Integrated Intensity	-0.024 ^c	0.981	-0.003
HISTOGRAM	Mean HU	-0.037 ^c	0.971	-0.004
	Variance HU	-0.257 ^c	0.797	-0.031
	Skewness HU	-0.855 ^b	0.392	-0.104
	Kurtosis HU	-1.124 ^b	0.261	-0.136
	Median HU	-0.104 ^b	0.917	-0.013
	Minimum Grey Level HU	-0.470 ^b	0.638	-0.057
	10thPercentileHU	-0.238 ^b	0.812	-0.029
	25thPercentileHU	-0.141 ^b	0.888	-0.017

GLC M	50thPercentileHU	-0.104 ^b	0.917	-0.013
	75thPercentileHU	-0.629 ^b	0.529	-0.076
	90thPercentileHU	-0.556 ^b	0.578	-0.067
	Standard Deviation HU	-0.409 ^c	0.682	-0.050
	Maximum Grey Level HU	-0.153 ^b	0.879	-0.019
	Mode HU	-1.100 ^c	0.271	-0.133
	Interquartile Range HU	-0.128 ^c	0.898	-0.016
	Range HU	-0.434 ^c	0.664	-0.053
	Mean Absolute Deviation HU	-0.122 ^c	0.903	-0.015
	Robust Mean Absolute Deviation HU	-0.208 ^b	0.835	-0.025
	Median Absolute Deviation HU	-0.134 ^c	0.893	-0.016
	Coefficient of Variation HU	-0.428 ^c	0.669	-0.052
	Quartile Coefficient of Dispersion HU	-0.165 ^c	0.869	-0.020
	EntropyLog10 HU	-0.391 ^c	0.696	-0.047
	EntropyLog2 HU	-0.391 ^c	0.696	-0.047
	Area Under Curve Csh HU	-0.037 ^c	0.971	-0.004
	Uniformity HU	-0.031 ^b	0.976	-0.004
	Root Mean Square HU	-0.458 ^c	0.647	-0.056
	Maximum Histogram Gradient HU	-0.354 ^c	0.723	-0.043
	Max Histogram Gradient Grey Level HU	-0.061 ^c	0.951	-0.007
	Min Histogram Gradient HU	-0.061 ^c	0.951	-0.007
	Min Histogram Gradient Grey Level HU	-0.770 ^b	0.441	-0.093
	Joint Maximum	-0.275 ^b	0.783	-0.033

	Joint Average	-0.043 ^c	0.966	-0.005
	Joint Variance	-0.361 ^c	0.718	-0.044
	Joint Entropy Log2	-0.483 ^c	0.629	-0.059
	Joint Entropy Log10	-0.483 ^c	0.629	-0.059
	Difference Average	-0.623 ^c	0.533	-0.076
	Difference Variance	-0.599 ^c	0.549	-0.073
	Difference Entropy	-0.483 ^b	0.629	-0.059
	Sum Average	-0.043 ^c	0.966	-0.005
	Sum Variance	-2.237 ^b	0.025	-0.271
	Sum Entropy	-3.592 ^b	0.000	-0.436
	Angular Second Moment	-0.263 ^b	0.793	-0.032
	Contrast	-0.788 ^c	0.431	-0.096
	Dissimilarity	-0.623 ^c	0.533	-0.076
	Inverse Difference	-0.251 ^b	0.802	-0.030
	Normalised Inverse Difference	-0.538 ^c	0.591	-0.065
	Inverse Difference Moment	-0.263 ^b	0.793	-0.032
	Normalised Inverse Difference Moment	-0.776 ^c	0.438	-0.094
	Inverse Variance	-5.690 ^c	0.000	-0.690
	Correlation	-0.116 ^b	0.908	-0.014
	Autocorrelation	0.000 ^d	1.000	0.000
	Cluster Tendency	-2.237 ^b	0.025	-0.271
	Cluster Shade	-0.623 ^c	0.533	-0.076
	Cluster Prominence	-2.707 ^b	0.007	-0.328
GLRLM	Short Runs Emphasis	-0.330 ^c	0.741	-0.040
	Long Runs Emphasis	-0.342 ^b	0.732	-0.041
	Low Grey Level Run Emphasis	-0.104 ^b	0.917	-0.013
	High Grey Level Run Emphasis	-0.031 ^b	0.976	-0.004

	Short Run Low Grey Level Emphasis	-0.251 ^b	0.802	-0.030
	Short Run High Grey Level Emphasis	-0.128 ^b	0.898	-0.016
	Long Run Low Grey Level Emphasis	-0.012 ^b	0.990	-0.001
	Long Run High Grey Level Emphasis	-0.287 ^c	0.774	-0.035
	Grey Level Non-Uniformity	-0.031 ^b	0.976	-0.004
	Run Length Non-Uniformity	-0.477 ^b	0.634	-0.058
	Run Percentage	-0.306 ^c	0.760	-0.037
NGTDM	Coarseness	-0.055 ^c	0.956	-0.007
	Contrast	-0.147 ^b	0.883	-0.018
	Busyness	-0.831 ^b	0.406	-0.101
	Complexity	-0.012 ^c	0.990	-0.001
	Strength	-1.350 ^b	0.177	-0.164
GLSZM	Small Zone Emphasis	-0.287 ^b	0.774	-0.035
	Large Zone Emphasis	-1.302 ^c	0.193	-0.158
	Low Gray Level Zone Emphasis	-0.037 ^c	0.971	-0.004
	High Gray Level Zone Emphasis	-0.220 ^b	0.826	-0.027
	Small Zone Low Grey Level Emphasis	-0.202 ^b	0.840	-0.024
	Small Zone High Grey Level Emphasis	-0.196 ^b	0.845	-0.024
	Large Zone Low Grey Level Emphasis	-1.308 ^c	0.191	-0.159
	Large Zone High Grey Level Emphasis	-1.314 ^c	0.189	-0.159

	Grey Level Non-Uniformity	-0.428 ^b	0.669	-0.052
	Normalised Grey Level Non-Uniformity	-0.367 ^b	0.714	-0.045
	Zone Size Non-Uniformity	-0.232 ^b	0.816	-0.028
	Normalised Zone Size Non-Uniformity	-0.257 ^b	0.797	-0.031
	Zone Percentage	-0.147 ^c	0.883	-0.018
	Grey Level Variance	-1.277 ^b	0.202	-0.155
	Zone Size Variance	-1.314 ^c	0.189	-0.159
	Zone Size Entropy	-0.208 ^c	0.835	-0.025

Table 5.2: Comparison of radiomic features between the ipsilateral and contralateral parotid, based on a bin width of 5 HU. Features which showed a significant difference between the parotid glands are highlighted in bold; significance level set at $p < 0.5$. ^b = based on negative ranks, ^c = based on positive ranks, ^d = the sum of negative ranks equals the sum of positive ranks.

Feature Family	Radiomic Feature	Z	p value	Effect Size
SHAPE	Volume	-0.383 ^b	0.701	-0.046
	Approximate Volume	-0.377 ^b	0.706	-0.045
	Voxels Counting	-0.377 ^b	0.706	-0.045
	Surface Area	-0.337 ^b	0.736	-0.040
	Surface To Volume Ratio	-0.407 ^c	0.684	-0.049
	Compacity	-0.050 ^c	0.960	-0.006
	Compactness1	-0.178 ^c	0.858	-0.021
	Compactness2	-0.184 ^c	0.854	-0.022
	Spherical Disproportion	-0.085 ^c	0.932	-0.010
	Sphericity	-0.167 ^c	0.868	-0.020
	Asphericity	-0.085 ^c	0.932	-0.010
	Centre Of Mass Shift	-1.091 ^c	0.275	-0.130
	Maximum3DDiameter	-0.102 ^b	0.918	-0.012
	Integrated Intensity	-0.881 ^c	0.378	-0.105
HISTOGRAM	Mean HU	-0.465 ^c	0.642	-0.056
	Variance HU	-0.465 ^c	0.642	-0.056
	Skewness HU	-0.489 ^b	0.625	-0.058
	Kurtosis HU	-0.892 ^b	0.372	-0.107
	Median HU	-0.026 ^b	0.979	-0.003
	Minimum Grey Level HU	-0.541 ^b	0.588	-0.065
	10thPercentileHU	-0.208 ^c	0.835	-0.025
	25thPercentileHU	-0.026 ^c	0.979	-0.003
	50thPercentileHU	-0.026 ^b	0.979	-0.003
	75thPercentileHU	-0.577 ^c	0.564	-0.069
	90thPercentileHU	-0.986 ^b	0.324	-0.118
	Standard Deviation HU	-0.530 ^c	0.596	-0.063

Maximum Grey Level HU	-0.196 ^b	0.845	-0.023
Mode HU	-0.705 ^b	0.481	-0.084
Interquartile Range HU	-0.929 ^c	0.353	-0.111
Range HU	-0.413 ^c	0.680	-0.049
Mean Absolute Deviation HU	-0.483 ^c	0.629	-0.058
Robust Mean Absolute Deviation HU	-0.143 ^c	0.886	-0.017
Median Absolute Deviation HU	-0.459 ^c	0.646	-0.055
Coefficient of Variation HU	-0.389 ^c	0.697	-0.046
Quartile Coefficient of Dispersion HU	-0.922 ^c	0.357	-0.110
EntropyLog10 HU	-0.477 ^c	0.633	-0.057
EntropyLog2HU	-0.477 ^c	0.633	-0.057
Area Under Curve Csh HU	-0.465 ^c	0.642	-0.056
Uniformity HU	-0.032 ^c	0.974	-0.004
Root Mean Square HU	-0.892 ^c	0.372	-0.107
Maximum Histogram Gradient HU	-0.044 ^c	0.965	-0.005
Max Histogram Gradient Grey Level HU	-0.658 ^c	0.510	-0.079
Min Histogram Gradient HU	-0.260 ^c	0.795	-0.031
Min Histogram Gradient Grey Level HU	-0.705 ^b	0.481	-0.084

GLCM	Joint Maximum	-0.202 ^c	0.840	-0.024
	Joint Average	-0.500 ^c	0.617	-0.060
	Joint Variance	-0.495 ^c	0.621	-0.059
	Joint Entropy Log2	-0.477 ^c	0.633	-0.057
	Joint Entropy Log10	-0.477 ^c	0.633	-0.057
	Difference Average	-0.606 ^c	0.545	-0.072
	Difference Variance	-0.746 ^c	0.456	-0.089
	Difference Entropy	-0.477 ^c	0.633	-0.057
	Sum Average	-0.500 ^c	0.617	-0.060
	Sum Variance	-1.389 ^b	0.165	-0.166
	Sum Entropy	-2.887^b	0.004	-0.345
	Angular Second Moment	-0.138 ^c	0.891	-0.016
	Contrast	-0.910 ^c	0.363	-0.109
	Dissimilarity	-0.606 ^c	0.545	-0.072
	Inverse Difference	-0.243 ^c	0.808	-0.029
	Normalised Inverse Difference	-0.980 ^c	0.327	-0.117
	Inverse Difference Moment	-0.284 ^c	0.777	-0.034
	Normalised Inverse Difference Moment	-0.916 ^c	0.360	-0.109
	Inverse Variance	-5.308^c	0.000	-0.634
	Correlation	-0.056 ^c	0.956	-0.007
	Autocorrelation	-0.500 ^c	0.617	-0.060
	Cluster Tendency	-1.389 ^b	0.165	-0.166
	Cluster Shade	-0.664 ^c	0.507	-0.079
	Cluster Prominence	-2.402 ^b	0.016	-0.287
GLRLM	Short Runs Emphasis	-0.143 ^c	0.886	-0.017
	Long Runs Emphasis	-0.272 ^c	0.786	-0.033
	Low Grey Level Run Emphasis	-0.056 ^b	0.956	-0.007

	High Grey Level Run Emphasis	-0.459 ^c	0.646	-0.055
	Short Run Low Grey Level Emphasis	-0.304 ^b	0.761	-0.036
	Short Run High Grey Level Emphasis	-0.489 ^c	0.625	-0.058
	Long Run Low Grey Level Emphasis	-0.035 ^b	0.972	-0.004
	Long Run High Grey Level Emphasis	-0.658 ^c	0.510	-0.079
	Grey Level Non-Uniformity	-0.214 ^b	0.831	-0.026
	Run Length Non-Uniformity	-0.342 ^b	0.732	-0.041
	Run Percentage	-0.120 ^c	0.905	-0.014
NGTDM	Coarseness	-0.547 ^c	0.584	-0.065
	Contrast	-0.108 ^c	0.914	-0.013
	Busyness	-0.348 ^b	0.728	-0.042
	Complexity	-0.401 ^c	0.689	-0.048
	Strength	-0.799 ^b	0.424	-0.095
GLSZM	Small Zone Emphasis	-0.407 ^c	0.684	-0.049
	Large Zone Emphasis	-0.928 ^c	0.354	-0.111
	Low Gray Level Zone Emphasis	-0.430 ^b	0.667	-0.051
	High Gray Level Zone Emphasis	-0.629 ^c	0.529	-0.075
	Small Zone Low Grey Level Emphasis	-0.500 ^c	0.617	-0.060
	Small Zone High Grey Level Emphasis	-0.114 ^b	0.909	-0.014
	Large Zone Low Grey Level Emphasis	-0.448 ^c	0.654	-0.054

Large Zone High Grey Level Emphasis	-0.613^c	0.540	-0.073
Grey Level Non-Uniformity	-0.424^b	0.671	-0.051
Normalised Grey Level Non-Uniformity	-0.149^c	0.881	-0.018
Zone Size Non-Uniformity	-0.167^b	0.868	-0.020
Normalised Zone Size Non-Uniformity	-0.407^c	0.684	-0.049
Zone Percentage	-0.331^c	0.741	-0.040
Grey Level Variance	-0.863^b	0.388	-0.103
Zone Size Variance	-0.565^c	0.572	-0.068
Zone Size Entropy	-0.319^c	0.750	-0.038

Table 5.3: Comparison of radiomic features between the ipsilateral and contralateral parotid glands, based on a bin width of 10 HU. Features which showed a significant difference between the parotid glands are highlighted in bold; significance level set at $p < 0.5$. ^b = based on negative ranks, ^c = based on positive ranks

5.2.5 Univariate Analysis of Covariates and their Association with Xerostomia and Sticky Saliva

This study involved generating logistic regression modules to predict acute xerostomia and sticky saliva using a combination of clinical/dosimetric variables, and radiomics features. Prior to building predictive models, the data were analysed using descriptive and univariate statistics to obtain an overview of the patient data collected.

To gain better insight of patients' baseline parotid function, chi-squared tests were carried out to examine the relationship between patient sex, smoking status, and alcohol consumption with baseline xerostomia or sticky saliva. The association between patient age and baseline xerostomia/sticky saliva was testing using Mann Whitney U tests, due to the non-normal distribution (Shapiro Wilks $p > 0.05$) and continuous nature of this variable.

The above association tests were also performed on each of the dependant variables, as well as the average D_{mean} of the parotid glands, measured during treatment follow up (i.e., xerostomia or sticky saliva in the acute, sub-acute, and late time points).

Irrespective of the results of these tests, based on the known association of these clinical and dosimetric variables with salivary flow rate, xerostomia, and sticky saliva, the clinical and dosimetric variables mentioned above were all included in the predictive models built (204-206).

HPV status was recorded for patients diagnosed with oropharyngeal cancer; however, given this cohort contained a range of head and neck subsites, HPV status is not always relevant and as such was not included in the predictive models, due to the heterogeneity of the patient group.

Xerostomia and sticky saliva were also captured at baseline and similar studies have included this as an independent variable in the generation of the predictive models for acute or late xerostomia (207-209). Araújo et al (210)., caution that this approach may introduce bias and as such, the association between baseline xerostomia/sticky saliva was tested against acute, sub-acute and late xerostomia/sticky saliva using the chi-squared

test. Irrespective of these results, baseline events were not used in the final generation of the predictive models.

5.2.6 Dimension Reduction and Feature Selection

Dimension reduction, or feature selection, is an important step when training predictive models to reduce the risk of overfitting. Although LIFEx extracts a small number of radiomic features comparative to other software programmes, it still produces high dimensional data in which there is potentially a higher number of features compared to observations (127). In addition, there is risk of multicollinearity among the extracted features. Feature reduction, can be handled in different ways; for example, cluster analysis, feature reproducibility analysis, or collinearity analysis (211). Collinearity analysis using variance inflation factors (VIF) and Spearman correlation was first performed to explore the presence and severity of multicollinearity. Features with a $VIF > 5$ or correlation coefficient of > 0.7 or < -0.7 were considered correlated, with the latter being mapped on a correlation matrix (Figure 5.1 and 5.2).

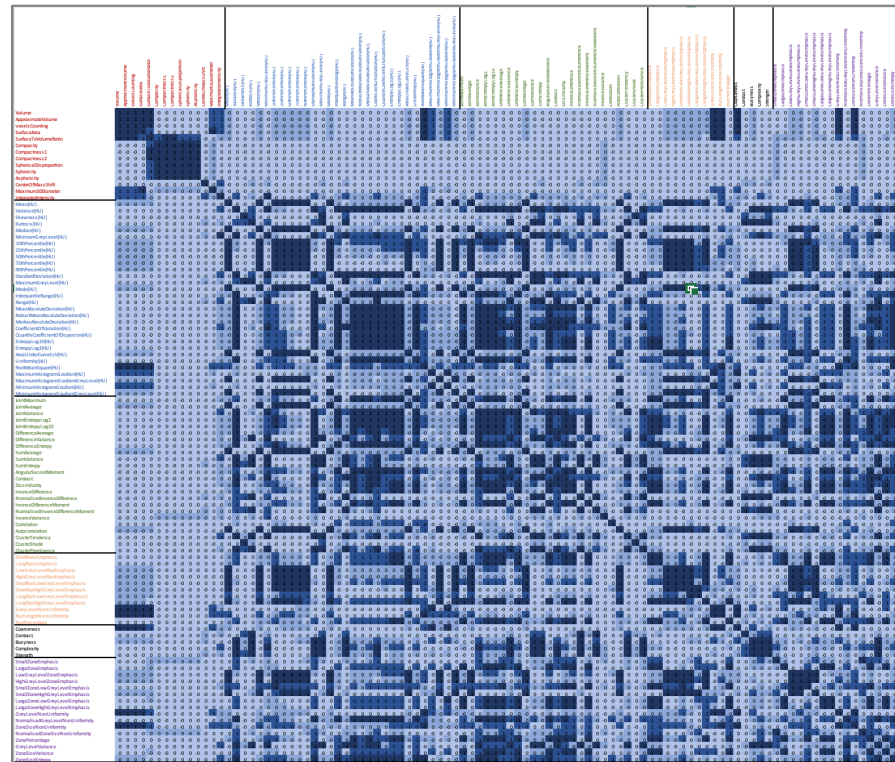


Figure 5.1 Spearman correlation matrix assessing collinearity of radiomic features extracted based on a bin width of 5 HU. Features with correlation coefficient of <-0.7 or >0.7 were considered correlated as indicated by the darkest blue.

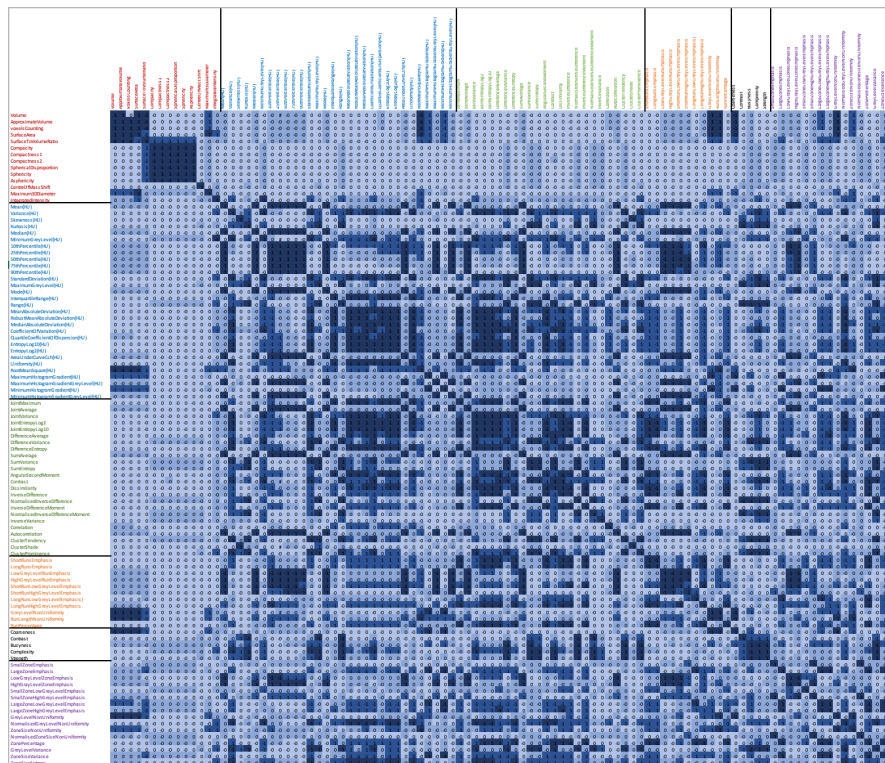


Figure 5.2 Spearman correlation matrix assessing collinearity of radiomic features extracted based on a bin width of 10 HU. Features with correlation coefficient of <-0.7 or >0.7 were considered correlated as indicated by the darkest blue.

Following this, to remove the presence of multicollinearity, principal component analysis (PCA) was performed. PCA eliminates collinearity between features by transforming them into a set of uncorrelated features with minimal information loss.

Firstly, the assumptions for conducting PCA were considered. These are:

1. Variables are continuous in nature.
 - a. This is the case for radiomic feature values.
2. Features have been standardised, due to sensitivity of the scale between the features.
 - a. This was handled through z-score normalisation of the features.
3. There is a linear relationship between the features.

- a. This was explored through the generation of scatterplots of randomly selected feature pairs.
- 4. There is sufficient sampling adequacy.
 - a. Although the number of cases relative to the number of observations. (i.e., radiomic features extracted) is low, PCA has been used in other radiomic studies using LIFEx with modest sample sizes (212, 213).

PCA works by reducing a large set of individual variables down to a set of principal components which can then be included in the predictive model. Critical to the process is first deciding on how many principal components to retain. This was based on a combination of scree plot analysis, which is a visual representation of the variation, and the Eigenvalue cut-off rule, which represents the total amount of variance which can be explained by each given component (Figures 5.3 and 5.4). Principal components with an Eigenvalue >1.000 were retained.

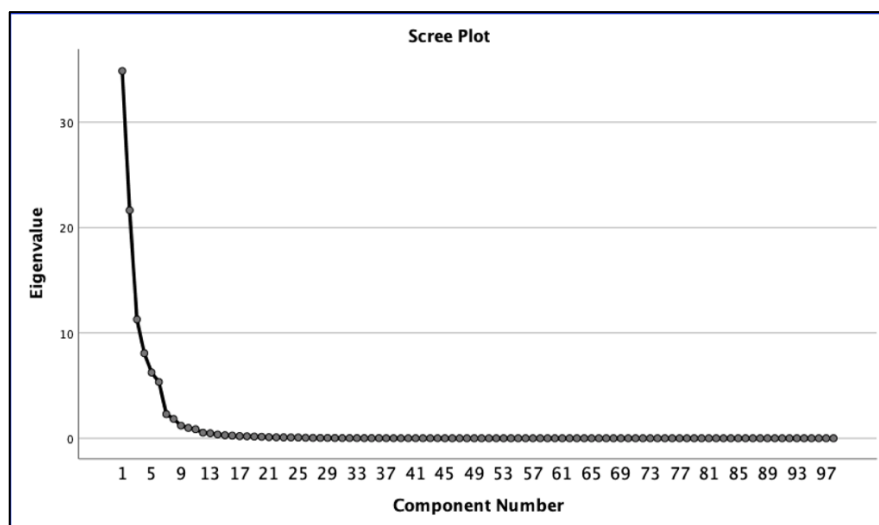


Figure 5.3: Scree plot of principal component analysis derived from radiomic features extracted based on a bin width of 5 HU.

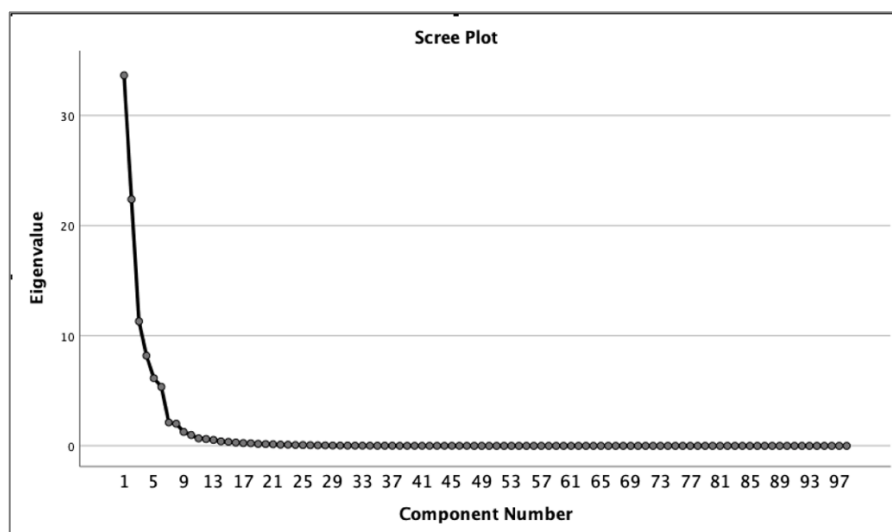


Figure 5.4: Scree plot of principal component analysis derived from radiomic features extracted based on bin width of 10.

This process identified nine principal components, explaining 94.68% and 94.29% of the total variation across the features for the analysis based on bin size of 5 HU and 10 HU respectively. To identify which radiomic features these principal components were most likely represented by the newly generated variables, the loading value of the radiomic features in the rotated component matrix were ranked, with the feature having the highest loading per component identified. These nine labelled, uncorrelated variables based on PCA were then included in the subsequent predictive models (Table 5.4). The PCA analysis was repeated for features extracted based on a bin size of 5 HU and again 10 HU (Table 5.5).

Component	Radiomic Feature Labelled	Loading Value	Initial Eigenvalues	% of Variance	Cumulative %
1	HISTOGRAM 75th Percentile (HU) *	0.977	34.863	35.574	35.574
2	HISTOGRAM Robust Mean Absolute Deviation (HU) *	0.939	21.645	22.087	57.661
3	NGTDM Strength *	0.943	11.285	11.516	69.177
4	GLRLM Run Percentage	0.931	8.071	8.236	77.412
5	GLRLM Grey Level Non-Uniformity	0.906	6.242	6.369	83.782
6	SHAPE Asphericity	0.972	5.343	5.452	89.234
7	GLSZM Small Zone Low Grey Level Emphasis	0.535	2.295	2.342	91.576
8	GLCM Normalised Inverse Difference Moment *	0.626	1.836	1.874	93.45
9	SHAPE Centre of Mass Shift *	0.635	1.204	1.229	94.679

Table 5.4: Nine components from principal component analysis with an eigenvalue >1. Radiomic features have been identified and labelled based on the loading value of the components. Results are from the PCA based on radiomic features extracted with a bin width of 5 HU. * Indicates radiomic features which were also labelled in the PCA when the bin width was equal to 10 HU.

Component	Radiomic Feature Labelled	Loading Value	Initial Eigenvalues	% Of Variance	Cumulative %
1	HISTOGRAM 75 th Percentile (HU) *	0.973	33.645	19.394	34.331
2	HISTOGRAM Robust Mean Absolute Deviation (HU) *	0.940	22.386	18.966	57.174
3	GLRLM Long Runs Emphasis	0.954	11.301	14.713	68.706
4	NGTDM Strength *	0.930	8.190	14.527	77.063
5	SHAPE Approximate Volume	0.908	6.142	10.261	83.331
6	SHAPE Compactness1	0.961	5.350	7.785	88.790
7	GLCM Inverse Variance	0.697	2.119	2.997	90.953
8	GLCM Normalised Inverse Difference Moment *	0.683	2.008	2.417	93.001
9	SHAPE Centre of Mass Shift *	0.832	1.264	1.345	94.291

Table 5.5: Nine components from principal component analysis with an eigenvalue >1. Radiomic features have been identified and labelled based on the loading value of the components. Results are from the PCA based on radiomic features extracted with a bin width of 10 HU. * Indicates radiomic features which were also labelled in the PCA when the bin width was equal to 5 HU.

5.2.7 Predictive Modelling

Based on the data available, a series of multivariant logistic regression models were developed, in which the severity of xerostomia, or sticky saliva as a binary outcome were the dependant variables.

5.2.7.1 Testing for Assumptions

There are several assumptions that must be satisfied for logistic regression (214):

1. The response variable is binary.
 - a. This assumption was met through the dichotomisation of xerostomia and sticky saliva, as previously described.
2. The observations are independent.
 - a. This assumption was met as there was no repeated measure from the same individual. Although presence of xerostomia and sticky saliva was obtained on several occasions, only one response per patient was collected for inclusion in each of the individual regression analyses.
3. No multicollinearity exists amongst the independent (exploratory) variables.
 - a. Regarding the radiomic features, this was handled using dimension reduction based on PCA. To assess the impact of this PCA and to confirm the lack of multicollinearity present following dimension reduction, VIF values were obtained on the remaining nine principal components. In addition, collinearity was assessed in the same way for the clinical and dosimetric features to be included in the final regression models (Tables 5.6 and 5.7)
4. For *continuous* independent (exploratory) variables, there should be a linear relationship between each one and the logit of the dependant (outcome) variable. This applies to age, mean parotid dose, and each the radiomic features derived from the principal components; all of which are *continuous* independent variables.

- a. This assumption was tested using the Box-Tidwell transformation. In order to perform this, the natural log of each continuous independent variable was computed, first transforming all data to return only positive values. This assumption is then satisfied if the goodness of fit for the resultant linear assessment is $p > 0.05$.

	Dependant Variables					
	Combined Models		Radiomics Models		Clinical Models	
Independent Variable	Xerostomia	Sticky Saliva	Xerostomia	Sticky Saliva	Xerostomia	Sticky Saliva
PC1 HISTOGRAM 75thPercentile (HU)	1.326	1.320	1.121	1.122		
PC2 HISTOGRAM Robust Mean Absolute Deviation (HU)	1.266	1.218	1.062	1.054		
PC3 GLRLM Lon Run Emphasis	1.338	1.317	1.108	1.105		
PC4 NGTDM Strength	1.328	1.324	1.121	1.119		
PC5 SHAPE Approximate Volume	1.433	1.420	1.021	1.021		
PC6 SHAPE Compactness1	1.501	1.506	1.025	1.026		
PC7 GLCM Inverse Variance	1.709	1.712	1.012	1.013		

PC8 GLCM Normalised Inverse Difference Moment	1.792	1.755	1.109	1.102		
PC9 SHAPE Centre of Mass Shift	1.488	1.496	1.080	1.080		
Age	1.178	1.205			1.026	1.029
Sex	1.761	1.760			1.117	1.116
Alcohol	2.184	2.092			1.630	1.620
Smoking	2.029	1.880			1.708	1.700
Mean Parotid Dose	1.303	1.294			1.025	1.025

Table 5.6: Collinearity assessment based on Variance Inflation Factors (VIF). VIF<5 indicates an absence of collinearity and satisfies one of the assumptions of logistic regression. Results shown are based on principal components derived from radiomic features extracted with a bin size of 5 HU.

	Dependant Variables					
	Combined Models		Radiomics Models		Clinical Models	
Independent Variable	Xerostomia	Sticky Saliva	Xerostomia	Sticky Saliva	Xerostomia	Sticky Saliva
PC1 HISTOGRAM 75thPercentile HU	1.102	1.102	1.132	1.132		
PC2 HISTOGRAM Robust Mean Absolute Deviation HU	1.611	1.611	1.048	1.048		
PC3 NGTDM Strength	1.577	1.577	1.093	1.093		
PC4 GLRLM Run Percentage	1.635	1.635	1.067	1.067		
PC5 GLRLM Grey Level Non Uniformity	1.285	1.285	1.019	1.019		
PC6 SHAPE Asphericity	1.317	1.317	1.033	1.033		

PC7 GLSZM Small Zone Low Grey Level Emphasis	1.181	1.181	1.012	1.012		
PC8 GLCM Normalised Inverse Difference Moment	1.199	1.199	1.095	1.095		
PC9 SHAPE Centre of Mass Shift	1.222	1.222	1.028	1.028		
Age	1.322	1.322			1.011	1.011
Sex	1.127	1.127			1.157	1.157
Alcohol	1.047	1.047			1.331	1.331
Smoking	1.245	1.245			1.393	1.393
Mean Parotid Dose	1.126	1.126			1.048	1.048

Table 5.7: Collinearity assessment based on Variance Inflation Factors (VIF). VIF<5 indicates an absence of collinearity and satisfies one of the assumptions of logistic regression. Results shown are based on principal components derived from radiomic features extracted with a bin size of 10 HU.

5.2.8 Predictive Models Built

Firstly, a logistic regression model based on patient clinical and dosimetric data as the independent variables was created. The clinical variables included patient age, sex, smoking history, history of alcohol consumption. It was intended that T stage would also be included as a covariate in the model; however, this was removed due to the amount of missing data. The dosimetric independent variable was the paired mean parotid dose for each individual patient.

To test the potential added value of radiomics, a second logistic regression model was created which, in addition to the aforementioned clinical and dosimetric independent variables, also included the nine labelled principal components. Finally, to test the hypothesis that radiomics alone is sufficiently powerful in predicting xerostomia a third logistic regression model was created based exclusively on the nine radiomic features labelled based on principal components.

The above regression models were fitted based on the same patient cohort to predict both acute xerostomia and sticky saliva as the dependant variables, resulting in total of six prediction models. It was intended to also build models to predict xerostomia and sticky saliva based on the sub-acute and late outcomes; however, due to limited data available at these timepoints and the imbalance of toxicity experienced by patients, these models were not generated. This decision was made due to the likelihood of separation occurring; a phenomenon in which there is perfect prediction of the model outcome, commonly caused by small cohorts and low incidence of an outcome event (215).

The entire process, including PCA dimension reduction was repeated based on radiomic features extracted based on a bin width of 10, to assess the reproducibility of results under these different pre-processing conditions.

5.2.8.1 Assessing Model Performance

The predictive performance of each model was assessed based on balanced accuracy (i.e. (sensitivity + specificity)/2)), Nagelkerke R², and the area under the receiver operating characteristic curve (AUC-ROC) (216). Ideally, the models' performance would be confirmed on a separate validation data set. Unfortunately, due to data availability and the limited study cohort, neither internal nor external validation was possible for this work. The study as a whole was assessed using the Radiomics Quality Score (107).

All data analysis was performed on IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp. Missing data were handled using the listwise deletion method in which a patient is removed from the analysis in the case of one or more missing variables. A shortened version of the CheckList for the EvaluAtion of Radiomic Research (CLEAR-S) of the study methodologic quality items is displayed in Figure 5.5 (217).

Section	No.	Item	Yes	No	n/a
Method					
<i>Study design</i>	7	Adherence to guidelines or checklists (e.g., CLEAR checklist)	☐	☒	☐
	8	Ethical details (e.g., approval, consent, data protection)	☒	☐	☐
	9	Sample size calculation	☐	☒	☐
	10	Study nature (e.g., retrospective, prospective)	☒	☐	☐
	11	Eligibility criteria	☒	☐	☐
	12	Flowchart for technical pipeline	☒	☐	☐
<i>Data</i>	13	Data source (e.g., private, public)	☒	☐	☐
	14	Data overlap	☒	☐	☐
	15	Data split methodology	☐	☐	☒
	16	Imaging protocol (i.e., image acquisition and processing)	☒	☐	☐
	17	Definition of non-radiomic predictor variables	☒	☐	☐
	18	Definition of the reference standard (i.e., outcome variable)	☒	☐	☐
<i>Segmentation</i>	19	Segmentation strategy	☒	☐	☐
	20	Details of operators performing segmentation	☐	☒	☐
<i>Pre-processing</i>	21	Image pre-processing details	☒	☐	☐
	22	Resampling method and its parameters	☒	☐	☐
	23	Discretization method and its parameters	☒	☐	☐
	24	Image types (e.g., original, filtered, transformed)	☒	☐	☐
<i>Feature extraction</i>	25	Feature extraction method	☒	☐	☐
	26	Feature classes	☒	☐	☐
	27	Number of features	☒	☐	☐
	28	Default configuration statement for remaining parameters	☐	☐	☐
<i>Data preparation</i>	29	Handling of missing data	☒	☐	☐
	30	Details of class imbalance	☐	☐	☒
	31	Details of segmentation reliability analysis	☐	☒	☐

Section	No.	Item	Yes	No	n/a
	32	Feature scaling details (e.g., normalization, standardization)	☒	☐	☐
	33	Dimension reduction details	☒	☐	☐
<i>Modeling</i>	34	Algorithm details	☒	☐	☐
	35	Training and tuning details	☐	☐	☒
	36	Handling of confounders	☐	☒	☐
	37	Model selection strategy	☐	☐	☒
<i>Evaluation</i>	38	Testing technique (e.g., internal, external)	☐	☐	☒
	39	Performance metrics and rationale for choosing	☒	☐	☐
	40	Uncertainty evaluation and measures (e.g., confidence intervals)	☐	☒	☐
	41	Statistical performance comparison (e.g., DeLong's test)	☐	☐	☒
	42	Comparison with non-radiomic and combined methods	☒	☐	☐
	43	Interpretability and explainability methods	☒	☐	☐
Open Science					
<i>Data availability</i>	53	Sharing images along with segmentation data [n/e]	☐	☒	☐
	54	Sharing radiomic feature data	☐	☒	☐
<i>Code availability</i>	55	Sharing pre-processing scripts or settings	☐	☒	☐
	56	Sharing source code for modeling	☐	☐	☒
<i>Model availability</i>	57	Sharing final model files	☐	☒	☐
	58	Sharing a ready-to-use system [n/e]	☐	☐	☒

Figure 5.5: CheckList for the EvaluAtion of Radiomic Research (CLEAR-S) of the study methodologic quality items.

5.3 RESULTS

5.3.1 Patient Characteristics

Data from 98 patients were collected, representing 14% of patients enrolled on the REACH RCT (30, 32). 74% of patients were male and the median age was 61 years (range: 50-81 years). With respect to lifestyle choices, 29% never consumed alcohol and 15% never smoked tobacco. 37% of patients still actively smoke and 32% still consume alcohol. The majority (70%) of patients were diagnosed with oropharyngeal cancer, and of those patients 52% of their cancers were positive for HPV expression; however,

these data were missing for 12% of patients with oropharyngeal cancer. Complete patient and disease characteristics are shown in Table 5.8 and a crosstabulation of HPV status by history of smoking and alcohol consumption is shown in Table 5.9.

Total		N=98	100 (%)
Age	Mean	63	
	Median	61	
	Range	50-81	
	Missing Data	2	
Sex	M	73	74
	F	24	24
	Missing Data	2	2
Smoking Status	Never	15	15
	Stopped	44	45
	Active	36	37
	Missing Data	3	3
Alcohol Consumption	Never	28	29
	Stopped	33	34
	Active	31	32
	Missing Data	6	6
Site	Oropharynx	69	70
	Oral Cavity	8	8
	Hypopharynx	12	12
	Larynx	9	9
HPV Status	Positive	36/69	52
	Negative	25/69	36
	Missing Data	8	12
T Stage	T1	11	11
	T2	19	19
	T3	27	28
	T4, 4a and 4b	26	27
	Missing	13	13
N Stage	N0-1	27	27
	N2-3	67	68
	Missing	4	4

Table 5.8: Patient and disease characteristics for the full cohort of 98 patients.

		Smoking Status					Alcohol Consumption				
		Active	Stopped	Never	Missing data	Totals	Active	Stopped	Never	Missing data	Totals
HPV Status	Positive	7	15	13	1	36	6	7	21	2	36
	Negative	13	11	0	1	25	11	10	2	2	25
	Missing Data	1	6	0	1	8	3	3	1	1	8
	Totals	21	32	13	3	69	20	20	24	5	69

Table 5.9: Crosstabulation of HPV status by history of smoking and alcohol consumption for patients diagnosed with oropharyngeal cancer.

5.3.2 Treatment Details

5.3.2.1 Chemotherapy and Randomisation

As per the REACH clinical trial protocol, patients were first assessed if they were fit to receive cisplatin-based chemotherapy and were then randomised accordingly (30). For those fit to receive cisplatin, they were randomised (1:1) to either arm A (cisplatin–RT) or B (avelumab–cetuximab–RT). Patients who were deemed unfit to receive cisplatin-based chemotherapy were randomised (1:1) to either arm C (avelumab–cetuximab–RT) or arm D (cetuximab–RT), as depicted in Figure 5.6. Of the 98 patients made available for this study, 69% were deemed fit to receive cisplatin. The groups were balanced in that 47% received the intervention of avelumab combined with either cisplatin or cetuximab. The full distribution amongst the arms is shown in Table 5.10.

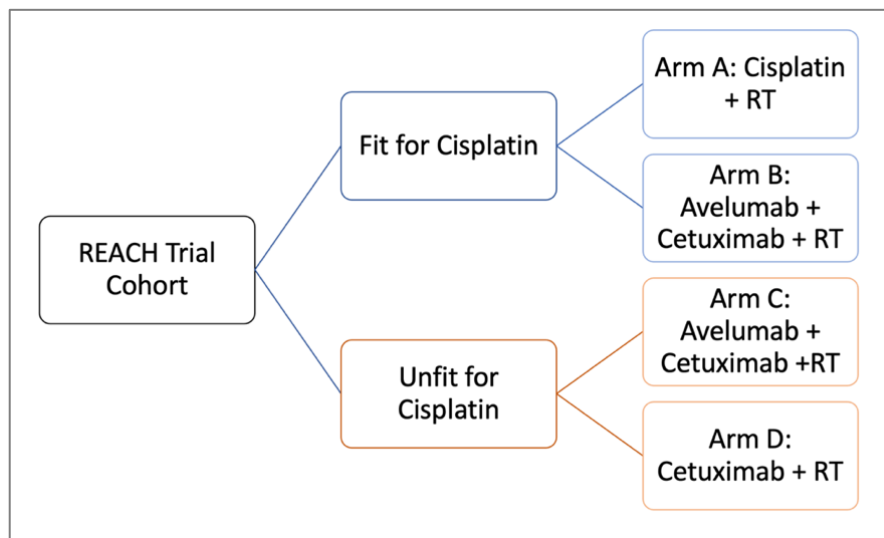


Figure 5.6: Trial design chart depicting randomisation of patients enrolled on the randomized trial of avelumab-cetuximab-radiotherapy versus standard of care in locally advanced squamous cell carcinoma of the head and neck (REACH).

Total		N=98	100 (%)
Trial Arm	A	33	34
	B	34	35
	C	12	12
	D	18	18
	Missing Data	1	1

Table 5.10: Randomisation of patients amongst trial arms for the full cohort of 98 patients. For patients fit to receive cisplatin, they were randomised (1:1) to either arm A (cisplatin–RT) or B (avelumab–cetuximab–RT). Patients who were deemed unfit to receive cisplatin-based chemotherapy were randomised (1:1) to either arm C (avelumab–cetuximab–RT) or arm D (cetuximab–RT).

5.3.2.2 Radiotherapy

As per the REACH trial protocol, patients were treated five days per week using an IMRT simultaneous integrated boost technique to a prescribed dose of 69.96 Gy in 33 fractions (2.12 Gy/fraction) to the high-risk disease and 52.8 Gy (1.6 Gy/fraction) to the low-risk region. Dose data were missing for 3% of patients. Of the remaining 96 patients, the average mean dose (D_{mean}) of the left and right parotid glands for these patients was 32.64 Gy $\pm$ 9.67 Gy and 35.68 Gy $\pm$ 10.25 Gy respectively. The left and right parotid glands' D_{mean} exceeded the established dose volume constraint of $D_{\text{mean}} < 25$ Gy in 81% and 88% of patients respectively (47).

5.3.3 Quality of Life Assessment: Incidence of Toxicity and Univariate Analysis

Figure 5.7 displays the incidence of baseline, acute, sub-acute, and late xerostomia, and sticky saliva.

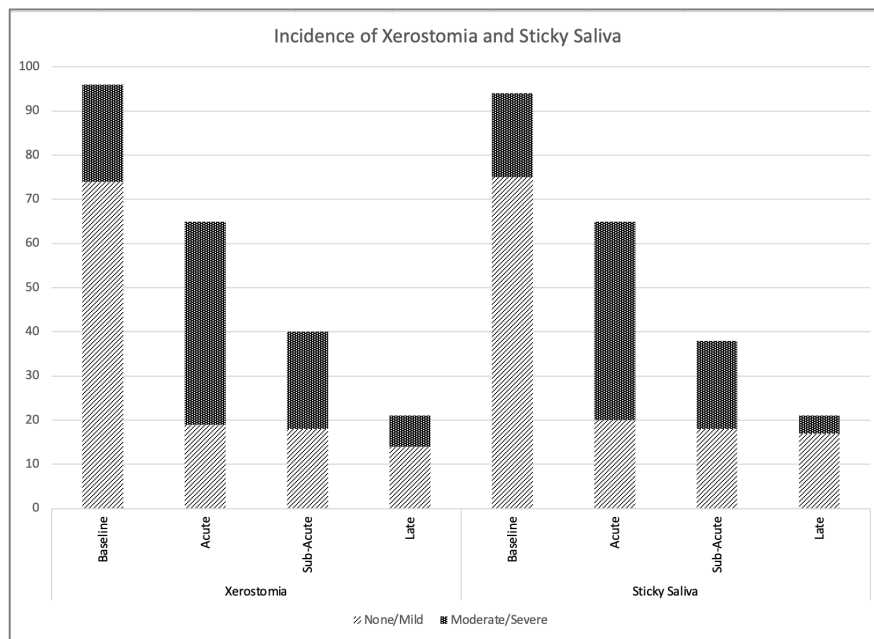


Figure 5.7: Incidence of baseline, acute, sub-acute and late xerostomia and sticky saliva dichotomised as none/mild and moderate/severe.

5.3.3.1 Baseline Assessment

At baseline, 74 out of the 96 patients (77%) with QoL data available presented with none or mild xerostomia. Two patients did not respond to question 42 at baseline, meaning their pre-treatment data regarding the presence of sticky saliva is missing. Seventy-five out of the remaining 94 patients (80%) reported none or mild severe sticky saliva at baseline.

Of the 22 patients (23%) who reported moderate/severe xerostomia at baseline:

- 7 were above the median age of 61 (missing age data n=2)
- 6 were female (missing sex data n=1)
- 10 still actively smoke, 9 previously smoked, 3 never smoked (Figure 5.8)
- 7 still actively consume alcohol, 10 no longer consume alcohol, 5 never consumed alcohol

Based on Chi-Squared and Mann Whitney U tests, there was no association between these variables, or mean dose to the parotid gland, and the prevalence of moderate to severe baseline xerostomia (Tables 5.11 and 5.12).

Of the 19 patients (20%) who reported moderate/severe sticky saliva at baseline:

- 5 were above the median age of 61 (missing age data n=2)
- 6 were female (missing sex data n=1)
- 8 still actively smoke, 8 previously smoked, 2 never smoked (missing smoking data n=1) (Figure 5.9)
- 8 still actively consume alcohol, 6 no longer consume alcohol, 4 never consumed alcohol (missing alcohol data n=1)

Of the above variables, only patient age was associated with an increased risk of presenting with moderate/severe sticky saliva at baseline ($p = 0.042$) (Table 5.11 and 5.12).

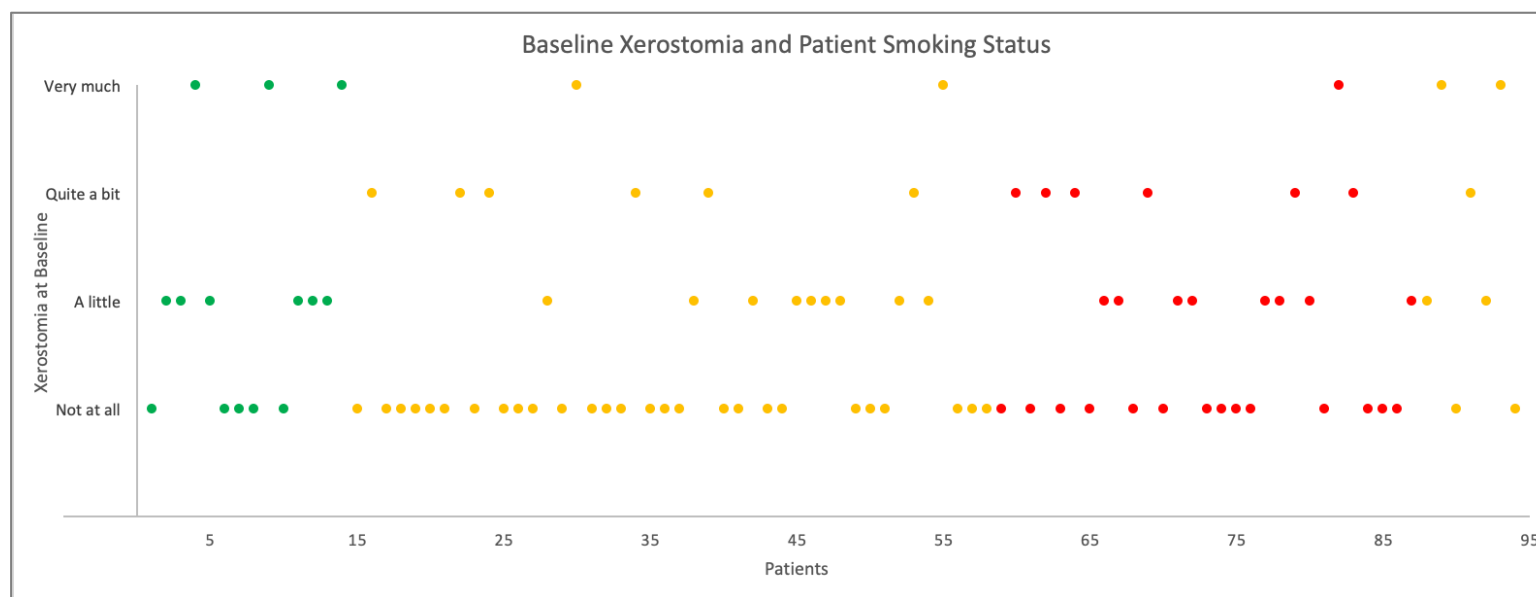


Figure 5.8: Scatterplot of patient smoking status and baseline xerostomia. Smoking status is coded: never smoked = green, stopped smoking = yellow, still actively smoking = red. Xerostomia scale taken from EORTC Quality of life questionnaire, Q42 “In the past week, have you had a dry mouth”

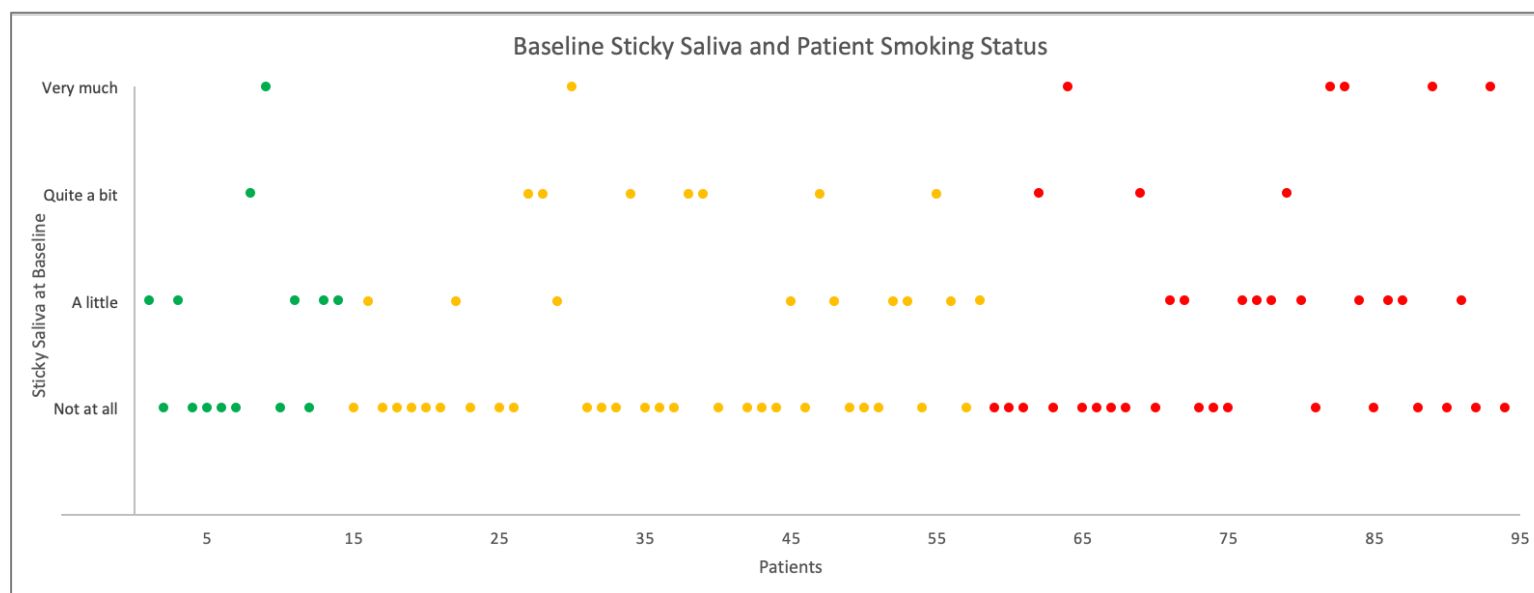


Figure 5.9: Scatterplot of patient smoking status and baseline sticky saliva. Smoking status is coded: never smoked = green, stopped smoking = yellow, still actively smoking = red. Xerostomia scale taken from EORTC Quality of life questionnaire, Q43 “In the past week, have you had a dry mouth.”

Xerostomia						Sticky Saliva					
		N	X ²	df	p value			N	X ²	df	p value
Baseline	Sex	92	3.653	2	0.161	Sex	92	5.065	4	0.281	
	Alcohol	93	1.856	2	0.395	Alcohol	93	1.688	4	0.793	
	Smoking	93	0.563	2	0.755	Smoking	93	1.876	4	0.759	
Acute	Sex	64	0.723	2	0.697	Sex	65	5.59	4	0.232	
	Alcohol	65	0.123	2	0.940	Alcohol	66	2.356	4	0.671	
	Smoking	65	1.522	2	0.467	Smoking	66	1.507	4	0.825	
Sub-Acute	Sex	39	2.198	1	0.138	Sex	37	0.579	1	0.447	
	Alcohol	40	3.416	2	0.181	Alcohol	38	10.614	2	0.050	
	Smoking	40	5.445	2	0.066	Smoking	38	4.738	2	0.094	
Late	Sex	21	0.875	1	0.350	Sex	21	2.085	1	0.149	
	Alcohol	21	0.824	2	0.662	Alcohol	21	3.896	2	0.143	
	Smoking	21	1.05	1	0.306	Smoking	21	0.031	1	0.861	

Table 5.11: Results of chi-Squared analysis testing the association between categorical clinical variables and baseline, acute, sub-acute, and late xerostomia/sticky saliva. N=number of observers included in analysis; df = degrees of freedom. a value for statistical significance is 0.05, and significant values are highlighted in bold.

	Xerostomia						Sticky Saliva		
	Variable	Event/Groups	N	Mean Ranks	U Statistic	p value	Mean Ranks	U Statistic	p value
Baseline	Age	None/Mild	73	49.23	567.000	0.127	48.79	453.000	0.042
		Moderate/Severe	20	38.85			34.67		
Acute	Age	None/Mild	19	35.50	389.500	0.493	34.80	414.000	0.608
		Moderate/Severe	46	31.97			32.20		
	Parotid Dose	None/Mild	17	36.26	301.500	0.201	28.08	334.500	0.340
		Moderate/Severe	45	29.70			32.90		
Sub-Acute	Age	None/Mild	18	21.64	177.500	0.577	20.08	169.500	0.759
		Moderate/Severe	22	19.57			18.98		
	Parotid Dose	None/Mild	17	19.00	170.000	0.803	17.00	136.000	0.419
		Moderate/Severe	21	19.90			19.84		
Late	Age	None/Mild	14	11.75	38.500	0.432	10.94	33.000	0.928
		Moderate/Severe	7	9.50			11.25		
	Parotid Dose	None/Mild	14	11.43	43.000	0.654	11.24	30.000	0.720
		Moderate/Severe	7	10.14			10.00		

Table 5.12: Comparison of patients with no or mild xerostomia and sticky saliva compared to patients with moderate or severe xerostomia and sticky saliva. Statistical significance is based on independent Mann Whitney U test and values in bold indicate a significant result ($p < 0.05$). NB: Parotid dose is the paired average mean dose of each patient's left and right parotid glands. The association analysis was not completed on parotid dose and baseline xerostomia/sticky saliva as the patient would not have received any dose at this point in time.

Based on chi-squared tests, no significant association between baseline toxicity scores and those captured during follow up was observed for xerostomia (Table 5.13). However, the presence of moderate/severe sticky saliva at baseline was associated with the same outcome in the acute (χ^2 (1) = 9.768, p = 0.045) and late setting (χ^2 (2) = 14.896, p = 0.001).

Baseline Xerostomia					Baseline Sticky Saliva				
	N	χ^2	df	p value		N	χ^2	df	p value
Acute Xerostomia	65	2.596	1	0.107	Acute Sticky Saliva	65	9.768	4	0.045
Sub-Acute Xerostomia	40	0.045	1	0.832	Sub-Acute Sticky Saliva	38	1.223	2	0.448
Late Xerostomia	21	0.276	1	0.599	Late Sticky Saliva	21	14.896	2	0.001

Table 5.13: Results of chi-Squared analysis testing the association between baseline xerostomia/sticky saliva and acute, sub-acute, and late xerostomia/sticky saliva. N=number of observers included in analysis; df = degrees of freedom. α value for statistical significance is 0.05, and significant values are highlighted in bold.

5.3.3.2 Acute Toxicity

Sixty-five patients had a questionnaire completed at least once within the first 5 months following baseline; 19 of which had two questionnaires completed in the acute period and two patients had three questionnaires completed in this phase. Of these 65 patients, 71% (n=46) patients experienced moderate/severe xerostomia. Data on acute sticky saliva were available for the same 65 patients, of which 45 (69%) experienced some moderate to severe amounts of sticky saliva. Patient age, sex, smoking or alcohol status, or mean dose to the parotid gland was not associated with acute moderate/severe xerostomia or sticky saliva. (Tables 5.11 and 5.12).

5.3.3.3 Sub-acute Toxicity

For the sub-acute phase, i.e., occurring between 5 to 17 months after baseline, toxicity data were available for 40 patients. Of these 40 patients, 16 patients had two questionnaires completed in during this period, and two patients had three questionnaires completed. This sub-acute period was the first-time toxicity had been captured after baseline for eight patients. Of the 40 patients with data captured during this period, 22 (55%) experienced moderate/severe xerostomia. Of these 40 patients, two did not answer the question related to sticky saliva; however, of the 38 patients who did, 20 of them (53%) reported experiencing moderate to severe toxicity. Of the variables considered in Tables 5.11 and 5.12, only history of alcohol consumption was significantly associated with an increased risk of developing moderate/severe sticky saliva in the sub-acute period following treatment ($\chi^2 (2) = 10.614, p = 0.05$)

5.3.3.4 Late Toxicity

Any data captured greater than 17 months after baseline were considered as late toxicity for this study. Twenty-one patients had data collected during this period, three of which had two questionnaires completed. For

one patient, the first time point that data was collected was 32 months following baseline.

Of these 21 patients, seven (33%) reported experiencing late moderate/severe xerostomia and 4 (19%) experiencing ongoing moderate/severe sticky saliva. Patient age, sex, smoking or alcohol status, or mean dose to the parotid gland were not associated with late xerostomia or sticky saliva (Tables 5.11 and 5.12).

5.3.4 Univariate Analysis of Radiomic Features

The association between each individual radiomic feature and acute, sub-acute, and late xerostomia/sticky saliva was examined using Mann Whitney U tests. Of the 100 radiomic features extracted with a bin width of 5 HU, 29% were significantly associated with at least one of the dependant variables (Table 5.14). Fifteen of these 29 features were associated with sub-acute xerostomia. GLCM_ Inverse Variance and GLCM_ Cluster Prominence were the only two features to be associated with more than one of the dependent variables. GLCM_ Inverse Variance being associated with both acute ($p = 0.045$) and late ($p = 0.048$) sticky saliva, and GLCM_ Cluster Prominence being associated with late xerostomia ($p = 0.017$) and sticky saliva ($p = 0.011$).

When the bin width was increased to 10 HU, 34% of the radiomic features were significantly associated with at least one of the dependant variables (Table 5.15).

Toxicity	Radiomic Feature	Event	N	Mean Rank	U Statistic	p value
Acute Xerostomia	GLCM Correlation	None/Mild	17	20.88	202	0.009
		Moderate/Severe	42	33.69		
Sub-Acute Xerostomia	SHAPE Maximum 3D Diameter	None/Mild	18	22.72	104	0.042
		Moderate/Severe	19	15.47		
	HISTOGRAM Mean (HU)	None/Mild	18	22.78	103	0.039
		Moderate/Severe	19	15.42		
	HISTOGRAM Median(HU)	None/Mild	18	22.81	102.5	0.037
		Moderate/Severe	19	15.39		
	HISTOGRAM 25thPercentile (HU)	None/Mild	18	22.58	106.5	0.049
		Moderate/Severe	19	15.61		
	HISTOGRAM 50thPercentile(HU)	None/Mild	18	22.81	102.5	0.037
		Moderate/Severe	19	15.39		
	HISTOGRAM 75thPercentile(HU)	None/Mild	18	22.72	104	0.041
		Moderate/Severe	19	15.47		
	HISTOGRAM Area Under Curve Csh (HU)	None/Mild	18	22.78	103	0.039
		Moderate/Severe	19	15.42		

	HISTOGRAM Minimum Histogram Gradient Grey Level(HU)	None/Mild	18	22.83	102	0.036
		Moderate/Severe	19	15.37		
	GLCM Joint Average	None/Mild	18	22.72	104	0.042
		Moderate/Severe	19	15.47		
	GLCM Sum Average	None/Mild	18	22.72	104	0.042
		Moderate/Severe	19	15.47		
	GLCM Auto-correlation	None/Mild	18	22.78	103	0.039
		Moderate/Severe	19	15.42		
	GLRLM Low Grey Level Run Emphasis	None/Mild	18	15.33	105	0.045
		Moderate/Severe	19	22.47		
	GLRLM High Grey Level Run Emphasis	None/Mild	18	22.72	104	0.042
		Moderate/Severe	19	15.47		
	GLRLM Short Run Low Grey Level Emphasis	None/Mild	18	15.17	102	0.036
		Moderate/Severe	19	22.63		
	GLRLM Short Run High Grey Level Emphasis	None/Mild	18	23.06	98	0.027
		Moderate/Severe	19	15.16		

Late Xerostomia	HISTOGRAM Variance(HU)	None/Mild Moderate/Severe	14 6	8.79 14.50	18	0.048
	HISTOGRAM Standard Deviation(HU)	None/Mild Moderate/Severe	14 6	8.79 14.50	18	0.048
	GLCM Joint Variance	None/Mild Moderate/Severe	14 6	8.79 14.50	18	0.048
	GLCM Difference Variance	None/Mild Moderate/Severe	14 6	8.79 14.50	18	0.048
	GLCM Cluster Prominence	None/Mild Moderate/Severe	14 6	8.43 15.33	13	0.017
	NGTDM Complexity	None/Mild Moderate/Severe	14 6	8.79 14.50	18	0.048
	GLSZM Grey Level Variance	None/Mild Moderate/Severe	14 6	8.57 15.00	15	0.026
	GLSZM Zone Size Entropy	None/Mild Moderate/Severe	14 6	8.29 15.67	11	0.011
Acute Sticky Saliva	GLCM Inverse Variance	None/Mild Moderate/Severe	18 41	33.56 28.44	305	0.045
Late Sticky Saliva	GLCM Difference Variance	None/Mild Moderate/Severe	16 4	9.19 15.75	11	0.047

GLCM Sum Variance	None/Mild Moderate/Severe	16 4	9.06 16.25	9	0.025
GLCM Sum Entropy	None/Mild Moderate/Severe	16 4	9.13 16.00	10	0.032
GLCM Inverse Variance	None/Mild Moderate/Severe	16 4	9.69 13.75	19	0.048
GLCM Cluster Tendency	None/Mild Moderate/Severe	16 4	9.06 16.25	9	0.025
GLCM Cluster Prominence	None/Mild Moderate/Severe	16 4	8.81 17.25	5	0.011
GLCM Strength	None/Mild Moderate/Severe	16 4	9.13 16.00	10	0.038

Table 5.14: Results of Mann Whitney U test examining the association between each radiomic feature extracted based on bin width of 5 HU and acute, sub-acute, and late xerostomia/sticky saliva. Analysis was conducted for each individual radiomic feature; however, only statistically significant results ($p < 0.05$) are displayed here.

Toxicity	Radiomic Feature	Event	N	Mean Rank	U Statistic	p value
Acute Xerostomia	GLCM Correlation	None/Mild	17	20.82	201	0.009
		Moderate/Severe	42	33.71		
Sub-Acute Xerostomia	GLCM Short Run High Grey Level Emphasis	None/Mild	17	36.94	239	0.048
		Moderate/Severe	42	27.19		
	SHAPE Maximum 3D Diameter	None/Mild	18	22.72	104	0.042
		Moderate/Severe	19	15.47		
	HISTOGRAM Mean (HU)	None/Mild	18	22.78	103	0.039
		Moderate/Severe	19	15.42		
	HISTOGRAM Median (HU)	None/Mild	18	22.67	105	0.044
		Moderate/Severe	19	15.53		
Sub-Acute Xerostomia	HISTOGRAM 25thPercentile (HU)	None/Mild	18	22.86	101.5	0.034
		Moderate/Severe	19	15.34		
	HISTOGRAM 50thPercentile (HU)	None/Mild	18	22.67	105	0.044
		Moderate/Severe	19	15.53		
	HISTOGRAM 90thPercentile (HU)	None/Mild	18	22.69	104.5	0.042
		Moderate/Severe	19	15.5		
Sub-Acute Xerostomia	HISTOGRAM Mode (HU)	None/Mild	18	22.64	105.5	0.046
		Moderate/Severe	19	15.55		

	HISTOGRAM Area Under Curve Csh (HU)	None/Mild Moderate/Severe	18 19	22.78 15.42	103	0.039
	HISTOGRAM Minimum Histogram Gradient Grey Level (HU)	None/Mild Moderate/Severe	18 19	22.86 15.34	101.5	0.034
	GLCM Joint Average	None/Mild Moderate/Severe	18 19	22.72 15.47	104	0.042
	GLCM Sum Average	None/Mild Moderate/Severe	18 19	22.72 15.47	104	0.042
	GLCM Autocorrelation	None/Mild Moderate/Severe	18 19	22.78 15.42	103	0.039
	GLRLM Low Grey Level Run Emphasis	None/Mild Moderate/Severe	18 19	15.39 22.42	106	0.048
	GLRLM High Grey Level Run Emphasis	None/Mild Moderate/Severe	18 19	22.72 15.47	104	0.042
	GLRLM Short Run Low Grey Level Emphasis	None/Mild Moderate/Severe	18 19	15.28 22.53	104	0.042
	HISTOGRAM Variance (HU)	None/Mild Moderate/Severe	14 6	8.79 14.5	18	0.048
Late Xerostomia		None/Mild	14	8.79	18	0.048

	HISTOGRAM Standard Deviation (HU)	Moderate/Severe	6	14.5		
	GLCM Joint Variance	None/Mild	14	8.79	18	0.048
		Moderate/Severe	6	14.5		
	GLCM Difference Variance	None/Mild	14	8.79	18	0.048
		Moderate/Severe	6	14.5		
	GLCM Cluster Prominence	None/Mild	14	8.43	13	0.017
		Moderate/Severe	6	15.33		
	GLSZM Normalised Grey Level Non-Uniformity	None/Mild	14	12.21	18	0.048
		Moderate/Severe	6	6.5		
	GLSZM Grey Level Variance	None/Mild	14	8.64	16	0.032
		Moderate/Severe	6	14.83		
	GLSZM Zone Size Entropy	None/Mild	14	8.29	11	0.011
		Moderate/Severe	6	15.67		
Acute Sticky Saliva	GLSZM Zone Percentage	None/Mild	18	37.17	240	0.034
		Moderate/Severe	41	26.85		
Late Sticky Saliva	GLCM Difference Variance	None/Mild	16	9.19	11	0.047
		Moderate/Severe	4	15.75		
	GLCM Sum Variance	None/Mild	16	9.06	9	0.025
		Moderate/Severe	4	16.25		

GLCM Sum Entropy	None/Mild Moderate/Severe	16 4	9.19 15.75	11	0.040
GLCM Inverse Variance	None/Mild Moderate/Severe	16 4	9.22 15.63	11.5	0.006
GLCM Cluster Tendency	None/Mild Moderate/Severe	16 4	9.06 16.25	9	0.025
GLCM Cluster Prominence	None/Mild Moderate/Severe	16 4	8.81 17.25	5	0.011
GLCM Strength	None/Mild Moderate/Severe	16 4	9.13 16	10	0.038
GLSZM Grey Level Variance	None/Mild Moderate/Severe	16 4	9.13 16	10	0.038

Table 5.15: Results of Mann Whitney U test examining the association between each radiomic feature and acute, sub-acute, and late xerostomia/sticky saliva. Analysis was conducted for each individual radiomic feature based on a bin width of 10; however, only statistically significant results ($p < 0.05$) are displayed here.

5.3.5 Predictive Models: Acute Xerostomia

The first logistic regression model was performed to ascertain the effects of patient age, sex, smoking status, history of alcohol consumption, and the average of the mean dose received by the left and right parotid glands on the likelihood a patient would experience moderate/severe xerostomia. This logistic regression model, based on 54 patients, was able to explain 6% of the variance in acute xerostomia (Table 5.16).

Patients with a history of smoking were nearly three times more likely to experience acute xerostomia (OR = 2.803, 95% CI = 0.0352 – 22.351, $p = 0.331$) compared to those who never smoked. The same impact of lifestyle choices could not be said for the impact of alcohol consumption (OR = 0.365, 95% CI = 0.066 – 2.025). The remaining covariates had minimal impact on the likelihood of the patients experiencing acute xerostomia.

As a comparison, a second regression model was built using solely radiomic features alone to predict the likelihood of moderate/severe acute xerostomia. Approximately 27% of the variation in acute xerostomia could be explained by the model, based on 59 patients (Table 5.16).

Although the radiomic feature GLRLM_LRE demonstrated a fourfold increased risk of experiencing acute xerostomia, this was not statistically significant (OR = 4.021, 95% CI = 0.497 - 32.555, $p = 0.192$). In fact, of the nine radiomic features included in this model, the Robust Mean Absolute Variation (Histogram_RMAD) feature from the Intensity Histogram family was the only feature that was statistically significant ($p = 0.046$). Increasing value of this feature indicated patients were twice as likely to experience acute xerostomia (OR = 2.432, 95% CI = 1.016 – 5.821).

Finally, to show the added value when combining radiomics with standard clinical variables, a third logistic regression analysis was conducted. Similar to the radiomics only model, again, an increase in GLRLM_LRE value increased the risk of acute xerostomia by a factor of four; however, this was still not considered statistically significant (OR = 4.478, 95% CI = 0.377 - 53.185, $p = 0.235$). In the combined model, Histogram_RMAD was the only

covariate that reached statistical significance ($p = 0.029$). Patients with an increase in this feature value were nearly four times more likely of developing acute xerostomia compared to those whose Histogram_RMAD remained stable or decreased (Table 5.16).

Acute Xerostomia								
Clinical Model				Radiomics Model			Combined Model	
Variables in model	B	OR (95% CI)	p value	B	OR (95% CI)	p value	B	OR (95% CI) p value
Age	0.026	1.026 (0.945 - 1.114)	0.542				0.056	1.057 (0.939 - 1.191) 0.358
Sex	0.305	1.357 (0.317 - 5.805)	0.680				1.661	5.264 (0.375 - 73.979) 0.218
Alcohol Status	-1.007	0.365 (0.066 - 2.025)	0.249				-0.653	0.52 (0.058 - 4.663) 0.559
Smoking Status	1.031	2.803 (0.352 - 22.351)	0.331				0.255	1.291 (0.062 - 27.008) 0.869
Mean Parotid Dose	-0.019	0.981 (0.873 - 1.102)	0.745				-0.042	0.959 (0.798 - 1.152) 0.653
HISTOGRAM 75thPercentile (HU)				-0.088	0.916 (0.428 - 1.958)	0.821	0.041	1.042 (0.424 - 2.559) 0.929
HISTOGRAM RMAD (HU)				0.889	2.432 (1.016 - 5.821)	0.046	1.285	3.614 (1.142 - 11.436) 0.029

GLRLM							
Long Runs Emphasis		1.392	4.021 (0.497 - 32.555)	0.192	1.499	4.478 (0.377 - 53.185)	0.235
NGTDM							
Strength		-0.141	0.868 (0.389 - 1.940)	0.730	-0.277	0.758 (0.331 - 1.736)	0.513
SHAPE							
Approximate Volume		0.187	1.205 (0.667 - 2.179)	0.537	-0.079	0.924 (0.385 - 2.218)	0.859
SHAPE							
Compactness1		-0.583	0.558 (0.260 - 1.199)	0.135	-0.861	0.423 (0.168 - 1.066)	0.068
GLCM							
Inverse Variance		-0.523	0.592 (0.234 - 1.502)	0.270	-0.304	0.738 (0.229 - 2.38)	0.611
GLCM							
NIDM		0.208	1.231 (0.511 - 2.964)	0.643	0.276	1.318 (0.473 - 3.672)	0.597
SHAPE							
Centre Of Mass Shift		-0.279	0.756 (0.362 - 1.579)	0.457	-0.353	0.703 (0.262 - 1.887)	0.484

Table 5.16: Clinical, radiomic and combined logistic regression models for the prediction of acute xerostomia. OR = Odds ratio. Significant values ($p < 0.05$) are highlighted in bold. Models were generated based on a bin width of 5 HU.

Regarding overall predictive power of these three models, relying on clinical data or dose to the parotids alone resulted in the poorest performing model with an AUC-ROC of 0.644, and a balanced accuracy of 53%. Analysing radiomic features exclusively achieved an AUC-ROC of 0.769 and a balanced accuracy of 61%. However, incorporating radiomic features with clinical and dosimetric variables increased the predictive power of the model to an AUC-ROC of 0.823 and balanced accuracy of 77%, indicating strong discriminatory performance in distinguishing between patients with and without acute xerostomia (Table 5.17).

		Clinical Model	Radiomics Model	Combined Model
Xerostomia	AUC	0.644	0.769	0.823
	Sensitivity	100	92.9	100
	Specificity	6.7	29.4	53.3
	Balanced			
	Accuracy	53.35	61.15	76.65
	-2 Log			
	Likelihood	61.279	58.300	46.887
	Nagelkerke R ²	0.066	0.274	0.355
	Patients included in model (n)	54	59	54
Sticky Saliva	AUC	0.669	0.775	0.901
	Sensitivity	94.7	95.1	94.7
	Specificity	37.5	38.9	62.5
	Balanced			
	Accuracy	66.1	67	78.6
	-2 Log			
	Likelihood	58.269	60.318	35.714
	Nagelkerke R ²	0.181	0.265	0.605
	Patients included in model (n)	54	59	54

Table 5.17: Assessment of the model's ability to predict acute xerostomia and sticky saliva. AUC-ROC = Area under the receiver operating characteristic curve. Models were generated based on a bin width of 5 HU.

For the models generated based on a bin width of 10 HU, some similar patterns were seen. The ability to predict acute xerostomia based on AUC-ROC was the best for the combined model; however, the difference between this and the AUC-ROC from the radiomics only model was marginal when compared (Table 5.18). The odds of developing xerostomia were increased with a high Intensity Histogram RMAD (HU) value for both the radiomics only model and the clinical models; however, this was only statistically significant for the combined model ($p=0.054$). No other covariates were significantly associated with an increased odds of developing acute xerostomia (Table 5.19).

		Clinical Model	Radiomics Model	Combined Model
Xerostomia	AUC	0.644	0.754	0.797
	Sensitivity	100	95.2	94.9
	Specificity	6.7	29.4	33.3
	Balanced			
	Accuracy	53.35	62.3	64.1
	-2 Log			
	Likelihood	61.297	59.02	48.323
	Nagelkerke R ²	0.066	0.260	0.360
	Patients included in model (n)	54	59	54
Sticky Saliva	AUC	0.669	0.734	0.911
	Sensitivity	94.7	90.2	92.1
	Specificity	37.5	16.7	68.8
	Balanced			
	Accuracy	66.1	53.45	80.45
	-2 Log			
	Likelihood	58.269	62.714	35.447
	Nagelkerke R ²	0.181	0.218	0.609
	Patients included in model (n)	54	59	54

Table 5.18: Assessment of the model's ability to predict acute xerostomia and sticky saliva. AUC-ROC = Area under the receiver operating characteristic curve. Models were generated based on a bin width of 10 HU.

Acute Xerostomia									
Clinical Model				Radiomics Model			Combined Model		
Variables in model	B	OR (95% CI)	p value	B	OR (95% CI)	p value	B	OR (95% CI)	p value
Age	0.026	1.026 (0.945 - 1.114)	0.542				0.044	1.045 (0.936 - 1.168)	0.433
Sex	0.305	1.357 (0.317 - 5.805)	0.680				1.464	4.323 (0.438 - 42.644)	0.210
Alcohol Status	-1.007	0.365 (0.066 - 2.025)	0.249				-0.711	0.491 (0.058 - 4.128)	0.513
Smoking Status	1.031	2.803 (0.352 - 22.351)	0.331				0.642	1.9 (0.104 - 34.715)	0.665
Mean Parotid Dose	-0.019	0.981 (0.873 - 1.102)	0.745				-0.019	0.981 (0.825 - 1.167)	0.827
HISTOGRAM 75thPercentile (HU)				-0.097	0.907 (0.427 - 1.927)	0.800	0.018	1.018 (0.417 - 2.488)	0.969
HISTOGRAM Robust Mean Absolute Deviation				0.734	2.083 (0.942 - 4.606)	0.070	0.992	2.696 (0.984 - 7.389)	0.054

NGTDM Strength	-0.164	0.849 (0.393 - 1.833)	0.677	-0.26	0.771 (0.346 - 1.722)	0.526
GLRLM Run Percentage	-1.456	0.233 (0.036 - 1.491)	0.124	-1.221	0.295 (0.046 - 1.908)	0.200
GLRLM Grey Level Non-Uniformity	0.129	1.137 (0.626 - 2.065)	0.673	-0.089	0.915 (0.395 - 2.117)	0.835
SHAPE Asphericity	-0.521	0.594 (0.286 - 1.234)	0.162	-0.791	0.454 (0.187 - 1.100)	0.080
GLSZM Small Zone Low Grey Level Emphasis	-0.278	0.757 (0.345 - 1.659)	0.487	-0.21	0.811 (0.295 - 2.232)	0.685
GLCM Normalised Inverse Difference Moment	0.093	1.098 (0.484 - 2.489)	0.823	0.064	1.066 (0.41 - 2.771)	0.895
SHAPE Centre of Mass Shift	0.263	1.301 (0.604 - 2.805)	0.501	0.287	1.333 (0.512 - 3.471)	0.556

Table 5.19: Clinical, radiomic and combined logistic regression models for the prediction of acute xerostomia. OR = Odds ratio. Significant values ($p < 0.05$) are highlighted in bold. Results are based on radiomic analysis based on bin width of 10.

5.3.6 Predictive Models: Acute Sticky Saliva

The same three models (clinical, radiomics, and combined) were generated using acute moderate/severe sticky saliva as the outcome variable.

The clinical model was able to explain 18% of the variation in acute sticky saliva experienced by the 54 patients in the model. Females were four times more likely to experience moderate/severe xerostomia compared to their male counterparts (OR = 4.162, 95% CI = 0.939 – 18.459, $p = 0.061$). Age was not shown to have an impact on the likelihood of patients' experiencing moderate/severe xerostomia, nor was mean dose received by the parotid glands. With respect to lifestyle factors, patients who had a history of smoking were more likely to experience worse acute xerostomia; however, the same could not be said for patients with a history of alcohol consumption (Table 5.20).

Acute Sticky Saliva								
Variables in model	Clinical Model			Radiomics Model			Combined Model	
	B	OR (95% CI)	p value	B	OR (95% CI)	p value	B	OR (95% CI) p value
Age	0.032	1.032 (0.949 - 1.123)	0.457				0.032	1.032 (0.915 - 1.164) 0.604
Sex	1.426	4.162 (0.939 - 18.459)	0.061				2.466	11.772 (0.291-476.909) 0.192
Alcohol Status	-0.823	0.439 (0.084 - 2.299)	0.330				-2.271	0.103 (0.005 - 2.087) 0.139
Smoking Status	0.676	1.965 (0.251 - 15.413)	0.520				3.060	21.334 (0.281-1618.584) 0.166
Mean Parotid Dose	0.118	1.125 (0.989 - 1.28)	0.073				0.395	1.485 (1.05 - 2.1) 0.025
HISTOGRAM 75thPercentile (HU)				0.377	1.458 (0.71 - 2.994)	0.304	0.497	1.643 (0.47 - 5.74) 0.436

HISTOGRAM RMAD (HU)	0.503	1.654 (0.781 - 3.50)	0.189	0.685	1.985 (0.674 - 5.847)	0.214
GLRLM	0.422	1.526 (0.717 - 3.247)	0.273	0.545	1.724 (0.645 - 4.608)	0.277
Long Runs Emphasis						
NGTDM	-0.008	0.992 (0.431 - 2.281)	0.985	-0.399	0.671 (0.246 - 1.834)	0.437
Strength						
SHAPE	0.266	1.305 (0.721 - 2.361)	0.379	-0.172	0.842 (0.267 - 2.658)	0.769
Approximate Volume						
SHAPE	-0.660	0.517 (0.243 - 1.098)	0.086	-2.151	0.116 (0.02 - 0.683)	0.017
Compactness1						
GLCM	-0.926	0.396 (0.158 - 0.996)	0.049	-1.050	0.35 (0.076 - 1.604)	0.177
Inverse Variance						
GLCM	-0.090	0.914 (0.384 - 2.174)	0.838	-1.039	0.354 (0.075 - 1.667)	0.189
NIDM						
SHAPE	-0.238	0.788 (0.414 - 1.502)	0.469	-0.949	0.387 (0.11 - 1.358)	0.138
Centre Of Mass Shift						

Table 5.20: Clinical, radiomic and combined logistic regression models for the prediction of acute sticky saliva. OR = Odds ratio. Significant values ($p < 0.05$) are highlighted in bold. Models were generated based on a bin width of 5 HU.

Regarding the radiomics only model, approximately 27% of the variation in sticky saliva was explained by the model. Within this model, the highest odds ratio was 1.654 for Histogram_RMAD; however, the increased likelihood of experiencing moderate/severe xerostomia based on an increase in this feature value was not statistically significant ($p = 0.189$). The only radiomic feature which was statistically significant was GLCM_Inverse Variance ($p = 0.049$). The odds ratio of 0.396 indicates that a decrease of this radiomics feature is associated with an increased risk of moderate/severe sticky saliva.

Of the three models created, the model combining clinical, dosimetric, and radiomic features was able to explain the highest amount (60%) of variation in acute sticky saliva. In this model, the mean parotid dose demonstrated a moderate positive association with the outcome ($OR = 1.485$, 95% CI = 1.05 – 2.10); a result that was statistically significant ($p = 0.025$). Although insignificant in the radiomics only models, a decrease in the Morphological_Compactness1 feature was significantly associated with increased risk of acute sticky saliva ($OR = 0.166$, 95% CI = 0.02 – 0.683, $p = 0.017$).

A similar pattern can be seen with the overall performance of these three models as with the models generated to predict acute xerostomia (Table 5.17). The model relying on clinical and mean dose to the parotid gland achieved an AUC-ROC of 0.669 and a balanced accuracy of 66%. Incorporating radiomic features into the model, increased the performance to an AUC-ROC of 0.901, and balanced accuracy of nearly 79%.

For the models built to predict sticky saliva based on a bin width of 10 HU, the AUC-ROC was almost identical to those reported for the models generated with a bin width of 5 HU (Table 5.18). The mean parotid dose demonstrated a statistically significant, positive association with an increased risk of developing sticky saliva in the combined model ($OR = 1.485$ 95% CI = 1.05 - 2.1, $p = 0.025$). For the same model, a reduction in Morphological_Aspphericity was associated with a significantly reduced risk of developing sticky saliva ($OR = 0.14$, 95% CI = 0.03 - 0.654, $p = 0.012$). This

radiomic feature, however, was not labelled as one of the principal components for the models generated with a bin width of 5 HU, so a direct comparison cannot be made. Finally, for the combined model based on a bin width of 10, females were at a significantly increased risk of developing acute sticky saliva ($p = 0.033$) (Table 5.21).

Acute Sticky Saliva									
	Clinical Model			Radiomics Model			Combined Model		
Variables in model	B	OR (95% CI)	p value	B	OR (95% CI)	p value	B	OR (95% CI)	p value
Age	0.293	1.032 (0.949 - 1.123)	0.457				0.033	1.034 (0.918 - 1.165)	0.584
Sex	0.349	4.162 (0.939 - 18.459)	0.061				4.087	59.543 (1.379 - 2571.16)	0.033
Alcohol Status	-0.028	0.439 (0.084 - 2.299)	0.330				-1.993	0.136 (0.009 - 1.983)	0.145
Smoking Status	-0.482	1.965 (0.251 - 15.413)	0.520				2.455	11.643 (0.319 - 424.795)	0.181
Mean Parotid Dose	0.208	1.125 (0.989 - 1.28)	0.073				0.32	1.377 (1.07 - 1.772)	0.013
HISTOGRAM 75thPercentile (HU)				0.293	1.34 (0.666 - 2.696)	0.412	0.067	1.069 (0.392 - 2.915)	0.897
HISTOGRAM Robust Mean Absolute Deviation (HU)				0.349	1.418 (0.719 2.795)	0.313	0.477	1.611 (0.641 - 4.051)	0.311

NGTDM Strength	-0.028	0.972 (0.461 - 2.052)	0.941	-0.306	0.736 (0.298 - 1.819)	0.507
GLRLM Run Percentage	-0.482	0.617 (0.255 - 1.493)	0.284	-0.46	0.631 (0.246 - 1.619)	0.338
GLRLM Grey Level Non-Uniformity	0.208	1.231 (0.698 - 2.172)	0.473	-0.211	0.809 (0.285 - 2.299)	0.691
SHAPE Asphericity	-0.569	0.566 (0.285 - 1.126)	0.105	-1.964	0.14 (0.03 - 0.654)	0.012
GLSZM Small Zone Low Grey Level Emphasis	-0.568	0.566 (0.269 - 1.191)	0.134	-0.504	0.604 (0.193 - 1.891)	0.387
GLCM Normalised Inverse Difference Moment	-0.146	0.864 (0.408 - 1.829)	0.702	-0.86	0.423 (0.129 - 1.387)	0.156
SHAPE Centre of Mass Shift	0.435	1.545 (0.691 - 3.455)	0.290	1.68	5.364 (0.796 - 36.168)	0.085

Table 5.21: Clinical, radiomic and combined logistic regression models for the prediction of acute sticky saliva. OR = Odds ratio. Significant values ($p < 0.05$) are highlighted in bold. Results are based on radiomic analysis based on bin width of 10.

5.4 DISCUSSION

5.4.1 Baseline Events

Xerostomia is a complex and subjective condition influenced by many intrinsic and extrinsic factors. Whilst sticky saliva differs from xerostomia, the two conditions are linked, and as such is likely to be influenced by the same variables. Twenty-three percent of patients included in this study presented with moderate/severe baseline xerostomia. Jiang et al., (218) reported just 4% of over 400 patients experienced grade 2 xerostomia at baseline based on the CTCAE scoring system (57). The mean age of their population was very similar at 59 years; however, no details were provided on smoking or alcohol consumption. Using the same QLQ-HN35 (143) questionnaire, 10% of the 178 patients examined by Beetz et al., (219) reported moderate to severe xerostomia, and 13% reported moderate to severe sticky saliva at baseline. Again, no details were provided on lifestyle factors likely to influence xerostomia. Of note however, the median age of their study population was also similar to ours at 61.0 years (range: 32 - 85 years). Age of the study population is important to consider as in a multivariate analysis of over 600 patients, Villa et al., (220) demonstrated the prevalence of xerostomia increases with increasing age based a population of dental clinical attendees. In their study, the odds ratio for developing xerostomia was 2.8 (95%CI = 1.6–4.9) for patients over the age of 61, compared to 0.9 (95%CI = 0.5–1.8) for those in the 36–50-year age bracket. These results were rationalised due to the likely increase in medication use with age; another covariate which demonstrated increasing odds (OR = 2.2, 95%CI = 1.5–3.4) (220). Although in this study logistic regression models were not trained to predict incidence at baseline, surprisingly a significant association between patient age and baseline xerostomia was not observed ($p=0.127$); however, there was an association with baseline sticky saliva ($p=0.042$). For this cohort, of the 22 patients who presented with baseline xerostomia, just one third of them were above the mean age of 61. Similarly, a quarter of the 19 patients presenting with

baseline sticky saliva were above 61 years old. As a comparison, the median age of the patient cohort in Villa et al., (220) was younger at just 47 years. This difference in the age of the study populations is because those from Villa et al., (220) were recruited from a general dental clinic. The patient cohort in Study 3 along with Jiang et al., (218) and Beetz et al., (219) all had a diagnosis of head and neck cancer; a disease in which age is an established risk factor (221).

In a specifically older population of 215 patients with a mean age of 80.2 ± 4.5 years, Morita et al., (222) also noted that females were significantly associated (OR = 2.608, 95% CI = 1.174–5.791) with xerostomia, compared to males. Results of Study 3 did not demonstrate an association between patient sex and either xerostomia ($p=0.161$) or sticky saliva ($p=0.281$). As with Villa et al., (220) their cohort was also based patients from a community dental clinical; however, another key difference between their cohort and ours is the country of origin. Mortia et al., (222) conducted their analysis based on data from Japanese patients, which are likely to have different behaviours in terms of smoking and alcohol consumption compared to the French cohort used in this study (223, 224).

Examining the patients with oropharyngeal cancer, it can be seen from the crosstabulation in Table 5.9 that for patients with HPV negative disease, the overwhelming majority had a history of smoking and alcohol consumption. This observation is not surprising and re-iterates the carcinogenetic impact of these lifestyle choices on cancer development (221). For the patients with HPV positive disease, the majority of patients did not have a history of consuming alcohol. This observation has also been seen in the US population, based on analysis of SEER data, and may also be reflective of smoking trends within the French population (225).

Although these exposure and behavioural factors were not considered by Morita et al., (222) they did acknowledge their likely effect on xerostomia. Despite this, a significant association for these lifestyle behaviours with either baseline xerostomia or sticky saliva was not observed in this study

(Table 5.11). In fact, the only significant association noted in this regard was between alcohol consumption and sub-acute sticky saliva ($p=0.05$).

5.4.2 Acute and Late Events

Twenty-three percent and 20% of patients in this study reported baseline moderate/severe xerostomia and sticky saliva respectively. This increased to 71% and 69% within 5 months of baseline, and then ultimately decreased to 33% and 19% 12 months later. Jensen et al., (226) published a systematic review of xerostomia as a result of cancer treatment and reported the absence of moderate/severe (i.e. grade 3 and 4) xerostomia at baseline; however within 6 months of treatment completion this rose to 7%, stabilising at 6% within 1-2 years following treatment. It is interesting to see the overall incidence of toxicity is lower compared to the current results when some studies included in the review used non-modulated techniques; however, given the left and right parotid glands' D_{mean} exceeded the dose volume constraint of 25 Gy in 81% and 88% of patients respectively for this study, this rate of toxicity is perhaps not surprising (47).

Another point to consider here is how xerostomia is defined and graded. Although Jensen et al., (226) do not provide a comprehensive list of toxicity scoring tools used in the studies, several systems are mentioned to quantify the impact of patients' quality of life. Of the 11 studies examining impact of xerostomia on QoL following IMRT, only one used the EORTC QLQ-HN35 questionnaire as was used in this study (143). However, the Jansen et al., (226) review is now over 10 years old, and it is likely the use of this questionnaire has increased over the years, particularly given the recent report of its cross cultural validity (227). van Dijk et al., (184) published their biomarker modelling study in 2017, in which they used the EORTC QLQ-HN35 questionnaire and dichotomised the outcomes as done here. They reported moderate to severe xerostomia and sticky saliva rates of 40% and 25% respectively, 12 months following treatment; results similar to this current study. Unfortunately, despite the mean contralateral parotid dose being associated with xerostomia and sticky saliva based on univariate analysis ($p<0.001$), they do not detail the actual mean dose received by the

parotid glands, preventing a more complete comparison with this work. Similarly, baseline xerostomia was selected as a covariate in their model, due to univariate association with the outcomes ($p < 0.001$), yet they do not explicitly state the incidence of baseline events in their population, again limiting comparison with this cohort.

It is plausible that there would be a link between the severity of baseline xerostomia or sticky saliva at the severity reported following treatment, i.e., patients with more severe symptoms at baseline would report persistent or worse events following treatment (210). This association was not observed for xerostomia in the cohort used here ($p = 0.107$); however, was observed for sticky saliva at both acute ($p = 0.045$) and late ($p = 0.001$) timepoints.

5.4.3 Predictive Models

Several studies have re-examined predictive models for xerostomia and sticky saliva in the era of radiomics and machine learning, with the majority of them focusing on late rather than acute events (116, 129, 184, 185, 226). The primary aim of this chapter was to improve prediction of acute xerostomia and sticky saliva through the inclusion of radiomic analysis of the parotid glands for patients with head and neck cancer. Irrespective of bin size, the inclusion of radiomics into the models improved the predictive power for both acute xerostomia and sticky saliva; a result which is consistent with Sheikh et al., (116, 129, 184) who also specifically examined acute toxicity.

5.4.3.1 Clinical Model

Relying on clinical data and mean dose to the parotids alone to predict acute xerostomia was almost the equivalent of random guessing with an AUC-ROC of 0.64, and a balanced accuracy of 53%. Independent variables in this model included patient age, sex, smoking and alcohol consumption, and mean dose to the parotid glands. In 2012, Beetz et al., (219) developed

logistic regression models to predict acute xerostomia and sticky saliva 6 months following the completion of treatment, and resulted in an almost identical AUC-ROC to the current study at 0.68 and 0.67 for xerostomia and sticky saliva respectively. As well as predicting acute xerostomia, their study also shared many similarities with this clinical model, including the mean age of their population (61 years), the use of modulated techniques, and the use of the EORTC QLQ-HN35 (143), with the Likert scale being dichotomised as none/mild or moderate/severe. Other, more recent, studies have achieved more promising results for models that also only include clinical and dosimetric variables. One such example is van Dijk et al., (184) whose logistic regression model based on clinical/dosimetric data achieved an AUC-ROC of 0.75 for predicting moderate/severe xerostomia 12 months following treatment. Their approach was comprehensive in terms of exploring several potential independent clinical variables, including patient age, sex, WHO stage, weight, length and Body Mass Index, tumour TNM stage, tumour location, treatment technique and the use of systemic treatment, as well as mean dose to the parotid and submandibular glands. Risk of producing a model which suffers from overfitting was mitigated through the use of L1 regularization (LASSO), and of all the variables explored, the only non-radiomic variables identified to be included in the xerostomia models were mean dose to the contralateral parotid, and tumour stage. Soares et al., (209) took a somewhat “leaner” approach in the number of covariates included in their predictive models for the presence of RTOG grade 1 or 2 xerostomia within 12 months of treatment. Based on clinical expertise alone, they decided to include age, gender, baseline xerostomia, mean dose to the ipsi- and contralateral parotid glands within the models, without prior univariate analysis for covariate selection. Their study did, however, take a more comprehensive approach to model selection, in which a series of models were trained based on different machine learning algorithms. Although the best performing model based on random forest achieved an AUC-ROC of 0.73, the logistic regression model only achieved an AUC-ROC of 0.47. These results are interesting, especially as the majority of studies investigating

machine learning methods for predicting toxicity still favour supervised methods using logistic regression, and derivatives of such (210). Furthermore, these results do not align with the conclusions drawn by Gabrys et al., (185) and highlighted by Araujo et al., (210) which state that in small clinical studies, such as this, simple logistic regression models perform just as well as more complex machine learning approaches. The current model was based on commonly used logistic regression, however also incorporated patient lifestyle habits not included by either Soares et al., (209) van Dijk et al., or Beetz et al., (219)

The inclusion of baseline toxicity into predictive models is a contentious issue, with Araujo et al., (210) going so far as suggesting its inclusion will introduce bias within the model, and it was for this reason baseline events were not included in the models generated in this study. Despite this, many studies do include baseline xerostomia as an independent covariate when predicting xerostomia and sticky saliva using modelling (184, 208, 209, 218, 219). As one example, Jiang et al., (218) observed a statistically significant positive relationship between baseline xerostomia moderate/severe acute xerostomia measured 3 months following treatment based on the CTCAE system ($\beta=0.55$, OR=1.74, $p<0.001$). Although Beetz et al., (219) included baseline xerostomia and also showed similar results ($\beta=0.72$, OR=2.05, $p=0.05$) when predicting xerostomia six months following treatment, they purposefully excluded patients with moderate/severe baseline xerostomia in their cohort, which reduced their total sample size by 10%. This approach could potentially impact on model generalizability if the training set does not reflect a typical population. This point is reinforced with the data in Study 3, considering 23% of patients presented with moderate/severe xerostomia prior to treatment.

Several authors also all included tumour stage in their models (184, 218, 219). This was initially considered for this study; however, T stage was missing in 13% of our patients and N stage in 4%. With this level of missing data and resultant listwise deletion, concerns were raised over the impact of predictive power when training models of such a small cohort.

Furthermore, this analysis was based on a cohort of patients with a range of head and neck cancers. Knowing that T stage of some tumours in the head and neck is dependent on tumour size, and others dependant on depth of invasion or extension into adjacent structures, this further supports the decision to purposefully exclude disease stage in the models.

Deciding on which covariates to include within a model is a necessary step in model training, and researchers commonly take a statistical approach to exploring the association between independent and dependant variables (228). Similar to Soares et al., (209) the clinical variables included in this model were based on clinical knowledge, despite the lack of statistical association shown in the Mann Whitney U and chi-squared tests (Tables 5.11 and 5.12). Although not common, this approach is reasonable when working with small sample sizes due to the limited statistical power, and may have less bias than if feature selection was based purely on univariate analysis (209, 210, 214).

5.4.3.2 Combined models

The predictive models trained by Beetz et al., (219) lacked any kind of quantitative imaging features. The authors deemed the models' performance overall as satisfactory based on R^2 Nagelkerke score of 0.13 and 0.17 for xerostomia and sticky saliva respectively, and postulated there are likely to be other prognostic factors that remain unidentified. Although Jiang et al., (218) took a more comprehensive dosimetric analysis examining dose received at an voxel level by the parotid glands, they still neglect to integrate imaging analysis which could reveal complex biological information. It is here that radiomic analysis may assist in unveiling these "unidentified" factors alluded to by Beetz et al., (219) that extend beyond dosimetric and clinical characteristics.

van Dijk et al., (184) Nardone et al., (129) and Sheikh et al., (116) all explored the potential predictive gain with the inclusion of radiomic features as imaging biomarkers for xerostomia, either in the acute or late setting. The

models by van Dijk et al., (184) and Nardone et al., (129) were trained to predict xerostomia 12 months post treatment, both demonstrating improvements with the inclusion of radiomics. Nardone et al., (129) was the only other study which specifically used LIFEx radiomics software to predict xerostomia. For their model which combined clinical, dosimetric and radiomic features, the AUC-ROC was 0.91 compared to 0.78 without the inclusion of radiomics. Feature selection was based on univariate analysis, and of the 39 radiomic features extracted, only GLCM_Correlation and GLRLM_RLNU were included in the logistic regression models. GLRLM_RLNU was not associated with either acute ($p = 0.116$), sub-acute ($p = 0.715$) or late ($p = 0.248$) xerostomia for this study. GLCM_Correlation was associated with acute xerostomia in this study ($p = 0.009$); however, not with sub-acute or late xerostomia. Despite its positive association with acute xerostomia, the loading value of this feature within the PCA for the data extracted with a bin width of 5 HU was only apparent in Principal Component 2 (PC2) and Principal Component (PC3) with values of 0.766 and 0.436 respectively. These values were overshadowed by Histogram_RMAD and GLRLM_LRE which both had the highest loading values for PC2 and PC3, indicating these features most strongly influenced those components, and it was these features which were subsequently labelled within the regression models for acute xerostomia. The lack of loading of these specific features was also the case when the bin width was increased to 10 HU. In this PCA, GLRLM_LRE only had a positive loading value in PC5 with a value of 0.862, which was just surpassed by a feature within the same family (GLRLM_GLNU) with a loading value of 0.908. For the PCA based on bin width of 10, GLCM_Correlation loaded into PC2 with a value of 0.762; however as was the case with a bin with of 5 HU, this fell short in comparison to Histogram_RMAD which was subsequently selected for inclusion in the models.

Interestingly, when considering the specific features that were labelled from the PCAs of both the 5 HU and 10 HU analyses, of the nine principal components labelled, five were common to both; they were Morphological_Centre of Mass Shift, Histogram_75th Percentile,

Histogram_RMAD, NGTDM_Strength, and GLCM_NIDM (Table 5.22). This indicates that these specific features are particularly robust to changes in image discretization. Despite these features being identified as representing the majority of variation amongst all the features within the PCAs, only Histogram_75th Percentile and NGTDM_Strength were statistically associated with at least one of the outcomes (Table 5.14). When exploring covariate selection for logistic regression Ranganathan et al., (228) make an argument for increasing the level of statistically significant association from $p < 0.05$ to $p < 0.10$, essentially casting a wider net for inclusion in the model. When this higher threshold is taken, Morphological_Centre of Mass Shift could justifiably also be included in the models investigating acute sticky saliva ($p = 0.084$). Despite the apparent lack of association of the labelled radiomic features derived from the PCA, it must be acknowledged that several features had similarly high loading values within the PCA rotated component matrix, yet it was the feature with the absolute maximum value that was selected for labelling.

With Nardone et al., (129) using the same software to this study it provides an opportunity for more direct comparison of methodologies which may be system specific. Unfortunately, beyond listing the features extracted, they failed to provide any additional details regarding image pre-processing, such as resampling or discretisation. Nevertheless, their study did also include a similar mix of head and neck cancers, a similar cohort size ($n = 78$), a similar age of population (median age = 63; range = 39–81 years), and gender ratio to patients in this cohort. Despite these similarities, known methodological differences were present. Namely, the scoring system used (CTCAE vs EORTC QLQ-HN35), timing of xerostomia assessment, the purposeful exclusion of patients who presented with baseline xerostomia, and the exclusion of patients whose parotid DVC was achieved in planning. Furthermore, whilst they examined both parotid glands, the gland per patient with the highest V30 was taken as the ROI for that case. Methodological differences such as these were also apparent in other studies (116, 184).

Predicting late xerostomia was not possible in this study due to the amount of missing data and imbalance of events, resulting in model separation (215). Shiekh et al., (116) attempted to predict acute xerostomia 3 months following treatment, and timepoint closer to ours did. Their study explored several models with combinations of clinical, dosimetric and radiomic features extracted from both MRI and CT images. Of the eight combination models they trained, unfortunately none included an amalgamation of CT radiomic features with clinical and dosimetric data as presented here. Regardless, the models with the highest AUC-ROC values, were for the “combined” models; however, differences were modest.

Research which directly repeats the work of others is arguably redundant. The value added to the existing body of evidence is not in simply replicating previous results, but rather testing outcomes under different methodologies, as was done here. The similarity of the overall performance of our radiomic models irrespective of methodological differences with previous studies, simply confirms the predictive potential of radiomics. In this study, the models integrating radiomics with standard clinical and dosimetric variables outperformed the standard clinical models, irrespective of bin size (Tables 5.17 and 5.18).

5.4.3.3 Radiomics Only Model

When the benefits of combining radiomic features with standard clinical and dosimetric variables into predictive models are seen, it raises the question of whether these imaging biomarkers could replace the need for collecting clinical or dosimetric data, and whether it is possible to rely solely on radiomic features as predictors of clinical outcomes. This hypothesis was tested in this study, along with others (116, 184). When relying on just the nine radiomic features extracted from a bin width of 5 HU and derived from the principal component analysis, the model achieved an AUC-ROC of 0.77 and a balanced accuracy of 61% for predicting acute xerostomia. The predictive power of such a model was almost identical when the outcome

was changed to acute sticky saliva (AUC-ROC = 0.78, balanced accuracy = 67%). These values also remained consistent when the analysis was derived from a bin with of 10 HU. Although not as positive as the combined model, these results do demonstrate the strong predictive power of radiomic features. Of the several predictive models investigated by Sheikh et al., (116) the model trained relying exclusively on radiomic features derived from CT images to predict acute xerostomia achieved an AUC-ROC of 0.69. This study was similar to ours in that their patient cohort, albeit it significantly larger than ours (n=216), also included patients from a mix of head and neck cancer diagnoses. Furthermore, the predictive outcome of acute xerostomia, was also measured at a similar timepoint as ours, i.e., at 3 months following treatment. However, several critical differences between their work and ours must be acknowledged. Primarily, the radiomic features included in the model were derived for the ipsilateral and contralateral parotid and submandibular glands separately using PyRadiomics. Of the seven radiomic features which were selected for inclusion in the models, four of them were extracted from the parotid glands, with two features from the ipsilateral gland and two from the contralateral gland. Although the decision to combine features extracted from the parotid glands was based on the statistical similarity of results of each of the glands, it is not common and does limit more direct comparison with other studies in the field. However, even if the ROIs were examined in the same fashion, differences in software and pre-processing techniques also impact of study comparisons. Interestingly, Shiekh et al.,(116) used a bin with of 25, which, along with results presented in this study, suggests that whilst the bin width does influence raw radiomic feature values (229), this influence is arguably inconsequential when feeding these data into logistic regression models. Sheikh et al., (116) and van Dijk et al., (184) both extracted radiomic features based on python scripting (PyRadiomics). Whilst Nardone et al., (129) also used LIFEx as was done here, they did not build a models based on radiomic features exclusively, nor did they detail specific details regarding pre-processing.

5.4.4 Limitations of Study 3

Discussing results of this study in the context of the evidence base is challenged by subtle methodological differences which may influence results. These differences may be technical (radiomic specific methodology), clinical (toxicity assessment and timing of outcomes), or statistical (feature reduction and selection). Regardless, comparisons and contrasts have been made which served to advance our collective understanding of radiomic analysis for the prediction of xerostomia and sticky saliva.

The main limitation of this study is related to the small sample size and missing data. These missing data was not only apparent at baseline, but also during follow-up impacting on the total number of patients included in the predictive models. Ideally when developing a predictive model, the cohort is split into a training set, a validation set, and a test set; or at a minimum a training and validation set. k-fold cross validation is a common approach in many radiomic studies (116, 129, 184, 218), and it was intended to use 5-fold cross-validation for this chapter. This is a process of internal validation where the data set is randomly split into five equal groups or folds. One-fold is set aside as the validation set, and the model is trained on the patients in the remaining four folds. The process is then repeated five times with each fold acting as the validation set once, resulting in five separate models. The average performance of these five models is taken to be a measure of the model's overall performance. Alternatively, the individual model with the highest performance can then be re-validated based on the full cohort (i.e., all five folds combined).

When the cohort for this study was split into five folds, it resulted in unacceptably low numbers in both the training and validation sets, particularly for late xerostomia and sticky saliva in which there were high levels of missing data. As a result, it was decided that due to the small cohort size, internal validation would not be possible for this study, and the models would be trained on the whole group combined. Whilst van Calster et al., (230) present some compelling arguments regarding the overreliance

of model validation as an indicator of clinical acceptability, not completing this process risks unknowingly producing a model which suffers from overfitting and provides no detail on the model's generalizability.

Despite these limitations, the RQS of this study was 36.11%. Based on a review of radiomic studies in which the salivary glands were the region of interest, Aringhieri et al., (231) calculated the RQS of 23 studies and reported a mean RQS for the 10 CT based studies of $27.70\% \pm 12.28\%$; indicating results presented here are aligned with existing literature.

5.5 CONCLUSION

Incorporating radiomic analysis of the parotid glands improved the prediction of acute xerostomia and sticky saliva for patients with head and neck cancer. These results need to be validated on an external dataset to assess the presence of overfitting of these models.

6.1 INTRODUCTION

Based on longitudinal imaging, several studies have observed that parotid glands vary both in size and position during the course of treatment, resulting in an increase in dose received and further compounding the damage to the gland (232-235). Whilst these studies have informed us of the need to explore adaptive radiotherapy based on dosimetric assessment, they lack consideration of clinical endpoints, such as xerostomia. In addition, much of these earlier studies were conducted in the pre-radiomics era, and as such these changes observed are limited to the volume of the region of interest (ROI) alone. By using radiomics, we can extract greater quantitative detail about how the parotid glands change in response to radiation and how these changes relate to radiation induced xerostomia.

Using radiomic analysis, this study aims to quantify the shape, intensity, and textural changes of the parotid gland that occur during a course of radiotherapy to the head and neck region. In doing so, this study will test the hypothesis that the variability in individual radiomic features is different between patients that develop xerostomia compared to those who do not.

6.2 MATERIALS AND METHODS

This study used irrevocably anonymous data from the open access Cancer Imaging Archive (140), and as such did not require ethical approval.

6.2.1 Patient Demographics, Treatment, and Toxicity

A total of 52 parotid glands from 26 patients (20 male and 6 female) were included in this study. Patients ranged from 43 to 87 years old, with a median age of 66 years. Prior to treatment commencing the median weight of the patients was 163lbs (range: 84lbs – 227lbs) and following the completion of treatment, the median weight loss was 9lbs (range: 35 to -9lbs). All patients received VMAT radiotherapy with 6MV to doses ranging from 60 Gy to 70 Gy over 30 to 35 fractions at 2 Gy to 2.2 Gy per fraction for a range of tumours in the head and neck region. Sixty two percent (n=16) of patients received chemotherapy with the majority receiving Cisplatin. Patient characteristics are detailed in Table 6.1.

Patient Details			Tumour Details		Treatment Details				RTOG Xerostomia Score	
Patient	Age	Sex	Cancer Staging	Tumour Location	Chemotherapy Agent	Radiotherapy Total Dose (Gy)	Number of Fractions	Dose/Fx (Gy)	Near End Treatment	3-6 months Post Treatment
1	64	M	IVA	Left base of the tongue and retromolar trigone	Cisplatin, Cetuximab	69.96	33	2.12	2	1
2	87	M	IVA	Left base of the tongue	Carboplatin, Cetuximab	69.96	33	2.12	1	2
3	66	M	IVA	Left base of the tongue	Erbitux	70	35	2.00	0	2
4	52	M	IVA	Left tonsil	Cisplatin	69.96	33	2.12	0	2
5	71	M	III	Right tonsil with ipsilateral nodal disease	Cisplatin	69.96	33	2.12	0	2
6	59	M	IVA	Anterior floor of the mouth	Cisplatin	65	30	2.15	0	2

7	59	F	III	Left lateral border of the tongue	None	60	30	2.00	1	0
8	87	M	II	Left lateral oral tongue/tongue base	None	60	30	2.00	0	1
9	55	F	IIB	Right oral tongue	Cisplatin	66	30	2.20	0	0
10	69	M	IVA	Left mandible	Cisplatin	66	30	2.20	0	0
11	71	M	IVA	Lymph nodes of head, face, and neck	Cisplatin	66	33	2.00	0	0
12	80	M	IVA	Left base of the tongue, both tonsils with extensive lymph node	Cetuximab	66	33	2.00	0	0
13	84	M	IIA	Base of tongue	None	70	35	2.00	0	2
14	67	M	IVA	Larynx	None	60	30	2.00	0	0
15	59	M	III	Larynx (right true vocal cord)	Cisplatin	69.96	33	2.12	0	1
16	43	F	III	Right tongue base	Cisplatin	69.96	33	2.12	2	0

17	71	M	III	Nasal cavity	None	60	30	2.00	0	0
18	66	M	IVA	Oropharynx	Cisplatin (with Mannitol)	69.96	33	2.12	2	0
19	61	F	IVA	Larynx	Cisplatin	66	33	2.00	0	0
20	65	M	III	Larynx- right TVC	Carboplatin	69.96	33	2.12	0	0
21	68	M	III	Occipital scalp	Cisplatin	60	30	2.00	0	0
22	64	M	I	Right tonsil	None	63.6	30	2.12	0	0
23	80	M	IVA	Left hard palate	None	60	30	2.00	0	0
24	51	M	III	Paranasal sinus	None	60	30	2.00	0	0
25	45	F	IIA	Oral tongue	None	60	30	2.00	2	0
26	56	F	III	Nasal cavity	None	60	30	2.00	0	0

Table 6.1: Patient characteristics, treatment details, and RTOG xerostomia scores. M = male, F = female, Gy = gray, fx = fraction, RTOG = Radiation Therapy Oncology Group

Acute toxicity scores based on RTOG Acute Toxicity Scale for xerostomia were reported at mid treatment (A1), “near end” (A2), and 3-6 months post treatment (Late). Where an RTOG score was not recorded in the patient notes, the clinical description available was used to assign a grade, as previously reported by Gabrys et al., (185). Where xerostomia was not mentioned in the clinical notes, an assumption was made that the patient was not experiencing xerostomia at that time point. For the purpose of this study, xerostomia incidence was dichotomised into either xerostomia ($G \geq 1$), or no xerostomia ($G=0$) at each time point.

6.2.2 CT Images and Radiomics Feature Extraction

CT scans of the 26 patients were acquired on a Siemens CT scanner with 120kVp, 240mA and images were reconstructed using a B30s convolution kernel. These CT scans were acquired at three separate time points; prior to the course of radiotherapy starting (CT1), mid-treatment (CT2), and following the completion of the course of treatment (CT3).

Radiomic analysis was conducted using LIFEx software, version 6, and HU were binned using a fixed interval of 5 HU from -569 to 1632 for CT1, and -1000 to 1968 for CT2 and CT3. The values used for intensity range were based on the minimum and maximum range present in the patients’ parotid glands at each time point. Prior to radiomic feature analysis, all images were resampled to a common voxel size of 1x1x1mm.

The feature extraction was run to report on first order features (shape and histogram). Shape features include sphericity, compacity, surface, volume (mL) and volume (number of voxels). First order features derived from histograms include: skewness, kurtosis, excess kurtosis, entropy log10, entropy log2 and energy. In addition, 32 texture features were extracted for analysis. Each radiomic feature was extracted on both the left and right parotid gland as the region of interest. The radiomic feature values obtained from both parotid glands were then averaged for each patient (144).

6.2.3 Statistical Analysis

All data were first tested for normality using the Jarque-Bera test. To assess the variation in radiomic feature values over the three time points (CT1, CT2, and CT3), a Friedmans test was performed. Post hoc analysis with Wilcoxon signed-rank tests was conducted between CT1 and CT2, CT2 and CT3 and, CT1 and CT3. Given multiple comparisons were made, to reduce the risk of a Type 1 error occurring, a Bonferroni correction was applied to each Wilcoxon signed-rank test in which the level of statistical significance was determined by dividing the standard threshold of 0.05 by the number of comparisons made (i.e. 3).

Following this, the absolute paired percentage difference of each radiomic feature between CT1 and CT2, CT2 and CT3, and CT1 and CT3 was calculated for each patient. To test the hypothesis that the absolute paired percentage difference of each individual radiomic feature of patients that did not develop xerostomia is different to that of patients that did develop xerostomia, a Mann Whitney U test was performed. Each of the Mann Whitney U tests were repeated to compare differences in radiomic feature values between CT1 and CT2, CT2 and CT3 and, CT1 and CT3 for patients with xerostomia, reported at mid treatment (A1), “near end” (A2), and 3-6 months post treatment (Late).

In addition, to investigate the impact of the size, shape, and compacity of the parotid gland on xerostomia, as a secondary analysis, the results are also described for the Shape features not using the *absolute* values, but rather preserving the sign of the difference (i.e., positive, or negative values).

All statistical analyses were performed in IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp.

6.3 RESULTS

6.3.1 Xerostomia Scores

Full details of RTOG xerostomia scores reported at mid-treatment (A1), near-end of treatment (A2) and 3-6 months post-treatment (Late) are shown in Table 6.2 and Figure 6.1. The majority of patients did not report any acute xerostomia at either mid-treatment (n=20, 77%) or near the end of treatment (n=23, 88%). Following the completion of treatment 65% (n=17) of patients reported no late xerostomia present 3-6 months post-treatment. In total, 38% (n=10) patients never experienced any xerostomia either during or following treatment.

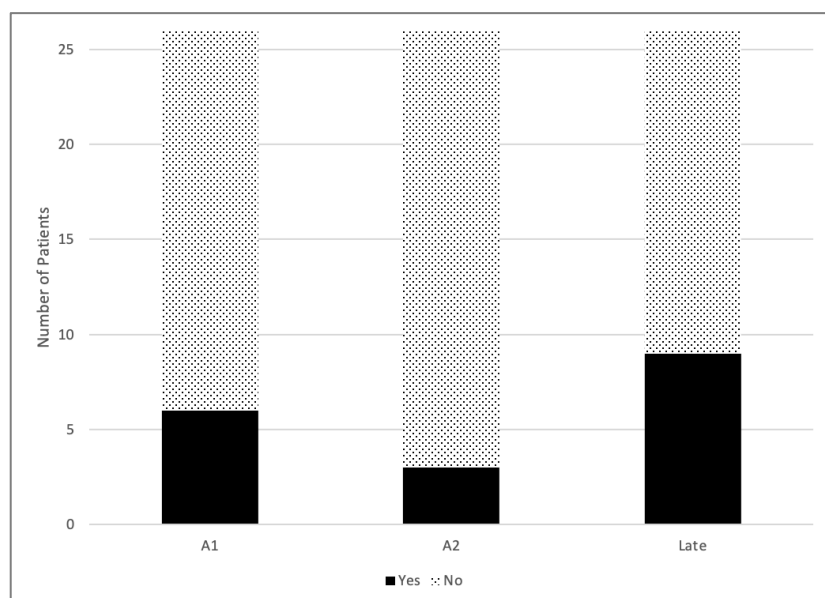


Figure 6.1: Incidence of xerostomia reported at three different timepoints; mid-treatment (A1), near-end of treatment (A2), and 3-6 months following the completion of treatment (Late). Total number of patients = 26.

	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Xerostomia	Not present	Slight dryness of mouth; good response on stimulation	Moderate dryness of mouth; poor response on stimulation	Complete dryness of mouth; no response on stimulation	Fibrosis	Death directly related to radiation effects
Mid Treatment	20	2	4	0	0	0
Near end Treatment	23	1	2			
3-6 months post treatment	17	3	6	1	0	0

Table 6.2: RTOG scores recorded at mid-treatment, near-end treatment, and 3-6 months post treatment. Total patients = 26

Figure 6.2 shows the mean absolute paired percentage difference of all radiomic features per patient. This scatterplot highlights there is higher variation in feature values observed during the first half of treatment (CT1 and CT2) compared with the second half of treatment (CT2 and CT3).

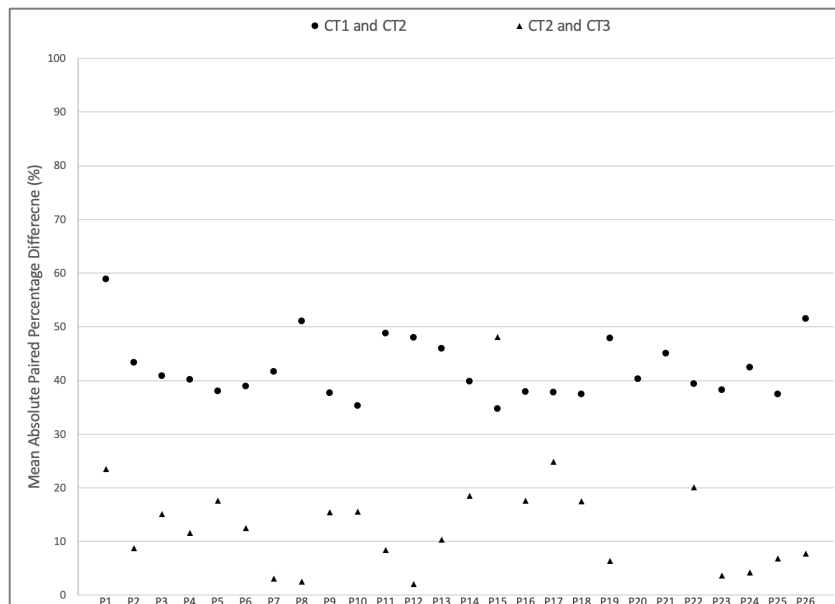


Figure 6.2: Mean absolute paired percentage difference of all radiomic features per patient. Radiomic features were extracted from CT imaging taken before the start of treatment (CT1), mid-way through treatment (CT2), and at end of treatment (CT3).

6.3.2 Radiomic Feature Variation

There was a statistically significant variation ($p < 0.05$) in 95% of radiomic feature across the three time points during the course of treatment (Table 6.3). The only two features which did not demonstrate significant variation in feature value were GLCM_Correlation ($X^2(2) = 1.462$, $p = 0.482$) and GLZLM_ZLNU ($X^2(2) = 2.846$, $p = 0.241$).

Radiomic Feature		Median	Chi-Square	df	p value
HISTO Skewness	CT 1	6.165	22.231	2	0.000
	CT 2	6.685			
	CT 3	6.542			
HISTO Kurtosis	CT 1	43.687	22.231	2	0.000
	CT 2	50.461			
	CT 3	49.240			
HISTO Excess Kurtosis	CT 1	40.687	22.231	2	0.000
	CT 2	47.461			
	CT 3	46.240			
HISTO Entropy_log10	CT 1	1.395	17.308	2	0.000
	CT 2	1.429			
	CT 3	1.446			
HISTO Entropy_log2	CT 1	4.633	17.308	2	0.000
	CT 2	4.748			
	CT 3	4.805			
HISTO Energy	CT 1	0.053	26.385	2	0.000
	CT 2	0.050			
	CT 3	0.048			
SHAPE Volume (mL)	CT 1	27.633	26.385	2	0.000
	CT 2	24.124			
	CT 3	21.494			
SHAPE Volume (vx)	CT 1	27632.500	26.385	2	0.000
	CT 2	24123.500			
	CT 3	21493.500			
SHAPE Sphericity	CT 1	0.608	4.846	2	0.089
	CT 2	0.603			
	CT 3	0.610			
SHAPE Surface (mm2)	CT 1	7425.539	23.615	2	0.000
	CT 2	6937.694			
	CT 3	6987.540			
SHAPE Compacity	CT 1	3.786	21.769	2	0.000
	CT 2	3.606			
	CT 3	3.376			
GLCM Homogeneity	CT 1	0.404	40.231	2	0.000
	CT 2	0.398			
	CT 3	3.376			
GLCM Energy	CT 1	0.006	22.462	2	0.000
	CT 2	0.005			
	CT 3	0.005			

GLCM Contrast	CT 1	31.093			
	CT 2	40.986	16.231	2	0.000
	CT 3	87.231			
GLCM Correlation	CT 1	0.705			
	CT 2	0.718	1.462	2	0.482
	CT 3	0.684			
GLCM Entropy_log10	CT 1	2.497			
	CT 2	2.563	13.000	2	0.002
	CT 3	2.576			
GLCM Entropy_log2	CT 1	8.295			
	CT 2	8.513	13.000	2	0.002
	CT 3	8.559			
GLCM Dissimilarity	CT 1	3.007			
	CT 2	3.231	16.692	2	0.000
	CT 3	3.709			
GLRLM SRE	CT 1	0.906			
	CT 2	0.908	12.538	2	0.002
	CT 3	0.916			
GLRLM LRE	CT 1	1.510			
	CT 2	1.500	9.538	2	0.008
	CT 3	1.471			
GLRLM LGRE	CT 1	0.000			
	CT 2	0.000	31.000	2	0.000
	CT 3	0.000			
GLRLM HGRE	CT 1	14080.859			
	CT 2	39272.384	40.923	2	0.000
	CT 3	39869.298			
GLRLM SRLGE	CT 1	0.000			
	CT 2	0.000	30.893	2	0.000
	CT 3	0.000			
GLRLM SRHGE	CT 1	12832.246			
	CT 2	35675.391	41.769	2	0.000
	CT 3	36492.111			
GLRLM LRLGE	CT 1	0.000			
	CT 2	0.000	32.077	2	0.000
	CT 3	0.000			
GLRLM LRHGE	CT 1	21939.164			
	CT 2	58280.673	39.692	2	0.000
	CT 3	58098.529			
GLRLM GLNU	CT 1	1269.682	30.692	2	0.000

	CT 2	1022.759			
	CT 3	898.950			
GLRLM RLNU	CT 1	18991.308			
	CT 2	16509.400	22.462	2	0.000
	CT 3	15263.73			
GLRLM RP	CT 1	0.873			
	CT 2	0.875	14.538	2	0.001
	CT 3	0.884			
NGLDM Coarseness	CT 1	0.001			
	CT 2	0.001	29.146	2	0.000
	CT 3	0.001			
NGLDM Contrast	CT 1	0.020			
	CT 2	0.024	13.231	2	0.001
	CT 3	0.030			
NGLDM Busyness	CT 1	0.092			
	CT 2	0.042	41.769	2	0.000
	CT 3	0.029			
GLZLM SZE	CT 1	0.624			
	CT 2	0.618	6.231	2	0.044
	CT 3	0.624			
GLZLM LZE	CT 1	2357.196			
	CT 2	1572.507	24.923	2	0.000
	CT 3	789.950			
GLZLM LGZE	CT 1	0.000			
	CT 2	0.000	31.462	2	0.000
	CT 3	0.000			
GLZLM HGZE	CT 1	14126.497			
	CT 2	39454.081	40.231	2	0.000
	CT 3	40690.115			
GLZLM SZLGE	CT 1	0.000			
	CT 2	0.000	31.462	2	0.000
	CT 3	0.000			
GLZLM SZHGE	CT 1	9029.778			
	CT 2	24334.813	41.769	2	0.000
	CT 3	25788.810			
GLZLM LZLGE	CT 1	0.1612			
	CT 2	0.0392	35.154	2	0.000
	CT 3	0.0308			
GLZLM LZHGE	CT 1	33910715.120			
	CT 2	61216315.880	19.000	2	0.000

	CT 3	30533933.930			
GLZLM GLNU	CT 1	166.929			
	CT 2	156.102	25.154	2	0.000
	CT 3	160.669			
GLZLM ZLNU	CT 1	1962.700			
	CT 2	1725.818	2.846	2	0.241
	CT 3	1826.903			
GLZLM ZP	CT 1	0.198			
	CT 2	0.207	12.538	2	0.002
	CT 3	0.2241			

Table 6.3: Friedman test results for radiomic feature values pre-treatment CT (CT 1), mid-treatment (CT 2), and post-treatment (CT 3). df = degrees of freedom. α value for statistical significance is 0.05.

Post hoc analysis using Wilcoxon signed-rank tests conducted with Bonferroni correction applied resulted in a significant level being set at 0.017, and complete results are presented in Table 6.4 with the descriptive statistics displayed in Table 6.5. Looking at the first half of treatment, i.e., comparing radiomic feature values between CT1 and CT2, it can be seen that 74% (n=32) of features were significantly different at CT2 compared to CT1 ($p<0.017$). When examining the second half of treatment (i.e., CT2 and CT3), this value dropped to 65% (n=28) ($p<0.017$).

Radiomic Feature	CT1 vs CT2			CT2 vs CT3			CT1 vs CT3		
	Z score	p value	Effect Size	Z score	p value	Effect Size	Z score	p value	Effect Size
HISTO_Skewness	-3.518 ^b	0.000	-0.690	-1.892b	0.058	-0.371	-2.908 ^b	0.004	-0.570
HISTO_Kurtosis	-3.314 ^b	0.001	-0.650	-2.045b	0.041	-0.401	-2.883 ^b	0.004	-0.565
HISTO_ExcessKurtosis	-3.314 ^b	0.001	-0.650	-2.045b	0.041	-0.401	-2.883 ^b	0.004	-0.565
HISTO_Entropy_log10	-3.035 ^b	0.002	-0.595	-2.933c	0.003	-0.575	-3.975 ^b	0.000	-0.780
HISTO_Entropy_log2	-3.035 ^b	0.002	-0.595	-2.933c	0.003	-0.575	-3.975 ^b	0.000	-0.780
HISTO_Energy	-4.051 ^c	0.000	-0.794	-1.943b	0.052	-0.381	-4.254 ^c	0.000	-0.834
SHAPE_Volume(mL)	-3.873 ^c	0.000	-0.760	-3.162b	0.002	-0.620	-4.254 ^c	0.000	-0.834
SHAPE_Volume(vx)	-3.873 ^c	0.000	-0.760	-3.162b	0.002	-0.620	-4.254 ^c	0.000	-0.834
SHAPE_Sphericity	-0.800 ^b	0.424	-0.157	-2.527b	0.012	-0.496	-1.333 ^c	0.182	-0.261
SHAPE_Surface(mm2)	-3.848 ^c	0.000	-0.755	-3.187b	0.001	-0.625	-4.203 ^c	0.000	-0.824
SHAPE_Compacity	-2.273 ^c	0.023	-0.446	-3.314b	0.001	-0.650	-3.848 ^c	0.000	-0.755
GLCM_Homogeneity	-1.816 ^c	0.069	-0.356	-3.010b	0.003	-0.590	-3.391 ^c	0.001	-0.665
GLCM_Energy	-3.721 ^c	0.000	-0.730	-2.426b	0.015	-0.476	-4.178 ^c	0.000	-0.819

GLCM_Contrast	-2.324 ^b	0.020	-0.456	-2.832c	0.005	-0.555	-4.280 ^b	0.000	-0.839
GLCM_Correlation	-0.114 ^c	0.909	-0.022	-.978b	0.328	-0.192	-0.876 ^c	0.381	-0.172
GLCM_Entropy_log10	-2.552 ^b	0.011	-0.500	-2.552c	0.011	-0.500	-3.772 ^b	0.000	-0.740
GLCM_Entropy_log2	-2.552 ^b	0.011	-0.500	-2.552c	0.011	-0.500	-3.772 ^b	0.000	-0.740
GLCM_Dissimilarity	-2.070 ^b	0.038	-0.406	-3.314c	0.001	-0.650	-4.102 ^b	0.000	-0.804
GLRLM_SRE	-1.892 ^b	0.058	-0.371	-3.060c	0.002	-0.600	-3.213 ^b	0.001	-0.630
GLRLM_LRE	-1.816 ^c	0.069	-0.356	-2.527b	0.012	-0.496	-2.629 ^c	0.009	-0.516
GLRLM_LGRE	-3.949 ^c	0.000	-0.774	-.927b	0.354	-0.182	-3.162 ^c	0.002	-0.620
GLRLM_HGRE	-4.457 ^b	0.000	-0.874	-2.172c	0.030	-0.426	-4.457 ^b	0.000	-0.874
GLRLM_SRLGE	-3.797 ^c	0.000	-0.745	-.296b	0.767	-0.058	-3.162 ^c	0.002	-0.620
GLRLM_SRHGE	-4.457 ^b	0.000	-0.874	-3.213c	0.001	-0.630	-4.457 ^b	0.000	-0.874
GLRLM_LRLGE	-4.432 ^c	0.000	-0.869	-1.765b	0.078	-0.346	-3.162 ^c	0.002	-0.620
GLRLM_LRHGE	-4.457 ^b	0.000	-0.874	-1.410b	0.159	-0.277	-4.457 ^b	0.000	-0.874
GLRLM_GLNU	-4.254 ^c	0.000	-0.834	-3.213b	0.001	-0.630	-4.381 ^c	0.000	-0.859
GLRLM_RLNU	-3.721 ^c	0.000	-0.730	-3.187b	0.001	-0.625	-4.203 ^c	0.000	-0.824
GLRLM_RP	-1.968 ^b	0.049	-0.386	-3.035c	0.002	-0.595	-3.035 ^b	0.002	-0.595
NGLDM_Coarseness	-4.130 ^b	0.000	-0.810	-1.968c	0.049	-0.386	-4.127 ^b	0.000	-0.809
NGLDM_Contrast	-2.552 ^b	0.011	-0.500	-2.603c	0.009	-0.510	-3.772 ^b	0.000	-0.740

NGLDM_Busyness	-4.457 ^c	0.000	-0.874	-2.603b	0.009	-0.510	-4.457 ^c	0.000	-0.874
GLZLM_SZE	-2.019 ^c	0.043	-0.396	-2.400c	0.016	-0.471	-0.851 ^b	0.395	-0.167
GLZLM_LZE	-3.543 ^c	0.000	-0.695	-2.959b	0.003	-0.580	-4.229 ^c	0.000	-0.829
GLZLM_LGZE	-3.797 ^c	0.000	-0.745	-1.384b	0.166	-0.271	-3.162 ^c	0.002	-0.620
GLZLM_HGZE	-4.457 ^b	0.000	-0.874	-2.197c	0.028	-0.431	-4.457 ^b	0.000	-0.874
GLZLM_SZLGE	-3.822 ^c	0.000	-0.750	-1.130b	0.258	-0.222	-3.162 ^c	0.002	-0.620
GLZLM_SZHGE	-4.457 ^b	0.000	-0.874	-2.603c	0.009	-0.510	-4.457 ^b	0.000	-0.874
GLZLM_LZLGE	-4.457 ^c	0.000	-0.874	-1.740b	0.082	-0.341	-3.162 ^c	0.002	-0.620
GLZLM_LZHGE	-3.467 ^b	0.001	-0.680	-2.857b	0.004	-0.560	-0.800 ^b	0.424	-0.157
GLZLM_GLNU	-2.959 ^c	0.003	-0.580	-2.756b	0.006	-0.540	-4.076 ^c	0.000	-0.799
GLZLM_ZLNU	-2.527 ^c	0.012	-0.496	-2.292c	0.770	-0.057	-2.121 ^c	0.034	-0.416
GLZLM_ZP	-1.206 ^b	0.228	-0.237	-3.060c	0.002	-0.600	-3.162 ^b	0.002	-0.620

Table 6.4: Results of statistical analysis (Wilcoxon signed rank test and effect size) on changes in radiomic feature values between CT taken pre-treatment (CT1) and CT taken at mid-treatment (CT2), CT taken at mid-treatment (CT2) and CT taken following the completion treatment (CT3) and, CT taken pre-treatment (CT1) and CT taken pre-treatment and following the completion treatment (CT3). b = based on negative rank; c = based on positive rank. Bonferroni correction was applied and $\alpha=0.017$ as level of statistical significance. All non-significant results are highlighted in bold.

Radiomic Feature		Mean	Std. Deviation	Median	IQR (1 st – 3 rd)
HISTO_Skewness	CT1	6.28	0.68	6.16	5.94 - 6.49
	CT2	6.73	0.52	6.68	6.46 - 6.87
	CT3	6.58	0.62	6.54	6.17 - 6.93
HISTO_Kurtosis	CT1	45.32	9.86	43.69	40.1 - 47.89
	CT2	51.46	7.57	50.46	47.47 - 53.60
	CT3	49.44	9.24	49.24	43.08 - 54.16
HISTO_ExcessKurtosis	CT1	42.32	9.87	40.69	37.1 - 44.89
	CT2	48.46	7.57	47.46	44.47 - 50.60
	CT3	46.44	9.24	46.24	40.08 - 51.16
HISTO_Entropy_log10	CT1	1.39	0.10	1.39	1.35 - 1.44
	CT2	1.42	0.07	1.43	1.37 - 1.46
	CT3	1.45	0.07	1.45	1.42 - 1.49
HISTO_Entropy_log2	CT1	4.61	0.31	4.63	4.5 - 4.79
	CT2	4.72	0.25	4.75	4.56 - 4.86
	CT3	4.83	0.22	4.80	4.7 - 4.95
HISTO_Energy	CT1	0.06	0.01	0.05	0.05 - 0.06

	CT2	0.05	0.01	0.05	0.04 - 0.05
	CT3	0.05	0.01	0.05	0.04 - 0.05
SHAPE_Volume_mL	CT1	27.55	7.94	27.63	21.69 - 32.61
	CT2	24.76	7.69	24.12	18.31 - 30.46
	CT3	22.89	7.07	21.49	17.12 - 28.94
SHAPE_Volume_vx	CT1	27545	7938.62	27632.50	21689.25 - 32606.75
	CT2	24763.92	7689.00	24123.50	18306.5 - 30456.25
	CT3	22890.92	7064.90	21493.50	17117.5 - 28941.25
SHAPE_Sphericity	CT1	0.61	0.06	0.61	0.56 - 0.65
	CT2	0.61	0.06	0.60	0.56 - 0.66
	CT3	0.6	0.06	0.61	0.53 - 0.64
SHAPE_Surface_CT1	CT1	7304.02	1666.28	7425.54	5634.23 - 8746.35
	CT2	6751.34	1568.66	6937.69	5303.57 - 7838.17
	CT3	6510.9	1530.73	6987.54	4956.61 - 7567.38
SHAPE_Compacity	CT1	3.74	0.46	3.79	3.4 - 4.10
	CT2	3.61	0.47	3.61	3.26 - 3.89
	CT3	3.47	0.50	3.38	3.17 - 3.86

GLCM_Homogeneity	CT1	0.41	0.04	0.40	0.38 - 0.42
	CT2	0.40	0.03	0.40	0.38 - 0.41
	CT3	0.39	0.03	0.38	0.37 - 0.40
GLCM_Energy	CT1	0.01	0.00	0.01	0.01 - 0.01
	CT2	0.01	0.00	0.01	0 - 0.01
	CT3	0.01	0.00	0.00	0 - 0.01
GLCM_Contrast	CT1	32.58	16.44	31.09	19.08 - 40.45
	CT2	52.19	34.64	40.99	23.72 - 85.23
	CT3	105.53	82.94	87.23	37.19 - 176.33
GLCM_Correlation	CT1	0.70	0.06	0.70	0.65 - 0.75
	CT2	0.70	0.07	0.72	0.67 - 0.75
	CT3	0.69	0.08	0.68	0.63 - 0.73
GLCM_Entropy_log10	CT1	2.49	0.17	2.50	2.43 - 2.60
	CT2	2.54	0.13	2.56	2.47 - 2.61
	CT3	2.59	0.12	2.58	2.54 - 2.66
GLCM_Entropy_log2	CT1	8.28	0.57	8.29	8.08 - 8.62
	CT2	8.45	0.44	8.51	8.21 - 8.68

	CT3	8.60	0.41	8.56	8.42 - 8.83
GLCM_Dissimilarity	CT1	3.11	0.55	3.01	2.75 - 3.42
	CT2	3.33	0.57	3.23	2.86 - 3.68
	CT3	3.93	0.83	3.71	3.18 - 4.59
GLRLM_SRE	CT1	0.90	0.02	0.91	0.9 - 0.91
	CT2	0.91	0.01	0.91	0.9 - 0.91
	CT3	0.91	0.01	0.92	0.91 - 0.92
GLRLM_LRE	CT1	1.54	0.12	1.51	1.46 - 1.56
	CT2	1.50	0.08	1.50	1.46 - 1.54
	CT3	1.49	0.11	1.47	1.44 - 1.52
GLRLM_LGRE	CT1	0.00	0.00	0.00	0.00 - 0.00
	CT2	0.00	0.00	0.00	0.00 - 0.00
	CT3	0.00	0.00	0.00	0.00 - 0.00
GLRLM_HGRE	CT1	14455.52	1361.58	14080.86	13372.26 - 15782.97
	CT2	39002.36	1781.66	39272.38	37526.04 - 40160.99
	CT3	39445.40	2254.14	39869.3	37054.78 - 40924.45
GLRLM_SRLGE	CT1	0.00	0.00	0.00	0.00 - 0.00

	CT2	0.00	0.00	0.00	0.00 – 0.00
	CT3	0.00	0.00	0.00	0.00 – 0.00
GLRLM_SRHGE	CT1	13051.51	1155.68	12832.25	12014.54 - 14196.55
	CT2	35422.69	1367.92	35675.39	34389.07 - 36313
	CT3	35997.99	1773.56	36492.11	34316.64 - 37196.82
GLRLM_LRLGE	CT1	0.00	0.00	0.00	0.00 – 0.00
	CT2	0.00	0.00	0.00	0.00 – 0.00
	CT3	0.01	0.03	0.00	0.00 – 0.00
GLRLM_LRHGE	CT1	22287.66	3305.03	21939.16	20140.12 - 23291.95
	CT2	58544.99	5178.79	58280.67	54959.34 - 61240.21
	CT3	58233.62	6970.03	58098.53	53912.55 - 60779.63
GLRLM_GLNU	CT1	1266.52	331.88	1269.68	1019.23 - 1412.48
	CT2	1036.15	302.58	1022.76	815.77 - 1262.11
	CT3	919.71	277.70	898.95	706.00 - 1145.13
GLRLM_RLNU_CT1	CT1	18860.56	5967.69	18991.31	14139.74 - 22527.90
	CT2	17182.31	5463.59	16509.4	12069.12 - 21381.29
	CT3	16086.68	5065.33	15263.73	11703.65 - 20422.11

GLRLM_RP	CT1	0.87	0.021	0.87	0.86 - 0.88
	CT2	0.88	0.015	0.88	0.87 - 0.88
	CT3	0.88	0.020	0.88	0.87 - 0.89
NGLDM_Coarseness	CT1	0.00	0.00	0.00	0.00 – 0.00
	CT2	0.00	0.00	0.00	0.00 – 0.00
	CT3	0.00	0.00	0.00	0.00 – 0.00
NGLDM_Contrast	CT1	0.02	0.01	0.02	0.02 - 0.02
	CT2	0.03	0.01	0.02	0.02 - 0.03
	CT3	0.04	0.04	0.03	0.02 - 0.04
NGLDM_Busyness	CT1	0.11	0.05	0.09	0.07 - 0.14
	CT2	0.05	0.03	0.04	0.03 - 0.06
	CT3	0.04	0.02	0.03	0.02 - 0.04
GLZLM_SIZE	CT1	0.62	0.02	0.62	0.61 - 0.64
	CT2	0.62	0.02	0.62	0.60 - 0.63
	CT3	0.63	0.02	0.62	0.61 - 0.64
GLZLM_LZE	CT1	3005.67	2144.78	2357.20	1871.90 - 3497.45
	CT2	1697.51	835.85	1572.51	1118.40 – 2606.00

	CT3	1271.04	1099.04	789.95	596.92 - 1724.29
GLZLM_LGZE	CT1	0.00	0.00	0.00	0.00 – 0.00
	CT2	0.00	0.00	0.00	0.00 – 0.00
	CT3	0.00	0.00	0.00	0.00 – 0.00
GLZLM_HGZE	CT1	14484.27	1495.93	14126.50	13134.18 - 15580.37
	CT2	39453.86	2006.21	39454.08	38377.55 - 15580.37
	CT3	40885.9	4167.00	40690.11	37257.21 - 43009.29
GLZLM_SZLGE	CT1	0.00	0.00	0.00	0.00 – 0.00
	CT2	0.00	0.00	0.00	0.00 – 0.00
	CT3	0.00	0.00	0.00	0.00 – 0.00
GLZLM_SZHGE	CT1	9134.31	1163.74	9029.78	8097.59 - 9855.06
	CT2	24587.13	1744.42	24334.81	23559.28 - 25992.03
	CT3	26421.96	4218.29	25788.81	23372.65 - 27849.21
GLZLM_LZLGE	CT1	0.21	0.14	0.16	0.12 - 0.25
	CT2	0.04	0.02	0.04	0.03 - 0.07
	CT3	8.70	33.58	0.03	0.02 - 0.05
GLZLM_LZHGE	CT1	44330314.60	34805323.30	33910715.11	25781189.10 – 49362500.00

	CT2	66096454.90	32720364.40	61216315.88	44725000.00 - 96503094.70
	CT3	50085320.60	46479986.90	30533933.93	22587500.00 - 61775000.00
GLZLM_GLNU	CT1	177.13	56.97	166.93	134.00 - 226.37
	CT2	163.87	50.71	156.10	120.69 - 211.81
	CT3	152.24	49.93	160.67	110.00 - 193.28
GLZLM_ZLNU	CT1	2088.14	838.75	1962.70	1392.00 - 2656.90
	CT2	1851.93	692.45	1725.82	1223.47 - 2391.11
	CT3	1896.20	691.18	1826.90	1370.19 - 2328.55
GLZLM_ZP	CT1	0.20	0.044	0.20	0.17 - 0.23
	CT2	0.21	0.041	0.21	0.18 - 0.23
	CT3	0.23	0.042	0.22	0.20 - 0.25

Table 6.5: Descriptive statistics of radiomic features extracted based on pre-treatment (CT1), mid-treatment (CT2), and at the completion of treatment (CT3). Radiomic features were extracted on the left and right parotid glands and values were averaged per patient.

6.3.3 Variation of Individual Radiomic Feature Values Associated with Xerostomia

On inspection of individual radiomic features it has been observed that the absolute paired percentage difference of some features varies between patients who experienced xerostomia, compared to those who did not (Tables 6.6, 6.7 and, 6.8).

Radiomic Feature	Event/Groups (Xerostomia at Mid Treatment)	N	CT1 and CT2			CT2 and CT3			CT1 and CT3		
			Mean Rank	p Value	U Statistic	Mean Rank	p Value	U Statistic	Mean Rank	p Value	U Statistic
HISTO_Skewness	No Acute Xerostomia	20	13.85	0.700	53.000	14.65	0.176	37.000	12.55	0.268	41.000
	Xerostomia	6	12.33			9.67			16.67		
HISTO_Kurtosis	No Acute Xerostomia	20	13.90	0.656	52.000	14.30	0.355	44.000	12.70	0.355	44.000
	Xerostomia	6	12.17			10.83			16.17		
HISTO_ExcessKurtosis	No Acute Xerostomia	20	13.90	0.656	52.000	14.30	0.355	44.000	12.70	0.355	44.000
	Xerostomia	6	12.17			10.83			16.17		
HISTO_Entropy_log10	No Acute Xerostomia	20	12.65	0.324	43.000	13.20	0.744	54.000	13.80	0.744	54.000
	Xerostomia	6	16.33			14.50			12.50		

HISTO_Entropy_log 2	No Acute Xerostomia	20	12.65	0.324	43.000	13.20	0.744	54.000	13.80	0.744	54.000
	Xerostomia	6	16.33			14.50			12.50		
HISTO_Energy	No Acute Xerostomia	20	12.75	0.387	45.000	14.20	0.421	46.000	13.70	0.836	56.000
	Xerostomia	6	16.00			11.17			12.83		
SHAPE_Volume_m L	No Acute Xerostomia	20	13.10	0.656	52.000	13.90	0.656	52.000	13.35	0.882	57.000
	Xerostomia	6	14.83			12.17			14.00		
SHAPE_Volume_vx	No Acute Xerostomia	20	13.10	0.656	52.000	13.90	0.656	52.000	13.35	0.882	57.000
	Xerostomia	6	14.83			12.17			14.00		
SHAPE_Sphericity	No Acute Xerostomia	20	13.55	0.976	59.000	14.95	0.083	31.000	13.40	0.929	58.000
	Xerostomia	6	13.33			8.67			13.83		
SHAPE_Surface	No Acute Xerostomia	20	13.45	0.976	59.000	13.65	0.882	57.000	13.40	0.929	58.000
	Xerostomia	6	13.67			13.00			13.83		
SHAPE_Compacity	No Acute Xerostomia	20	12.80	0.421	46.000	14.25	0.387	45.000	12.65	0.324	43.000
	Xerostomia	6	15.83			11.00			16.33		

GLCM_Homogeneity	No Acute Xerostomia	20	12.90	0.494	48.000	13.95	0.614	51.000	13.70	0.836	56.000
	Xerostomia	6	15.50			12.00			12.83		
GLCM_Energy	No Acute Xerostomia	20	12.90	0.494	48.000	14.05	0.533	49.000	13.80	0.744	54.000
	Xerostomia	6	15.50			11.67			12.50		
GLCM_Contrast	No Acute Xerostomia	20	13.25	0.790	55.000	14.90	0.095	32.000	14.40	0.295	42.000
	Xerostomia	6	14.33			8.83			10.50		
GLCM_Correlation	No Acute Xerostomia	20	15.25	0.033	25.000	14.10	0.494	48.000	14.70	0.157	36.000
	Xerostomia	6	7.67			11.50			9.50		
GLCM_Entropy_log 10	No Acute Xerostomia	20	12.65	0.324	43.000	13.20	0.744	54.000	14.20	0.421	46.000
	Xerostomia	6	16.33			14.50			11.17		
GLCM_Entropy_log 2	No Acute Xerostomia	20	12.65	0.324	43.000	13.20	0.744	54.000	14.20	0.421	46.000
	Xerostomia	6	16.33			14.50			11.17		
GLCM_Dissimilarity	No Acute Xerostomia	20	12.55	0.268	41.000	14.35	0.324	43.000	14.20	0.421	46.000

	Xerostomia	6	16.67			10.67			11.17			
GLRLM_SRE	No Acute Xerostomia	20	13.25	0.790	55.000	13.45	0.976	59.000	13.75	0.790	55.000	
	Xerostomia	6	14.33			13.67			12.67			
GLRLM_LRE	No Acute Xerostomia	20	13.45	0.976	59.000	12.90	0.494	48.000	13.50	1.000	60.000	
	Xerostomia	6	13.67			15.50			13.50			
GLRLM_LGRE	No Acute Xerostomia	20	13.25	0.790	55.000	12.30	0.157	36.000	14.00	0.573	50.000	
	Xerostomia	6	14.33			17.50			11.83			
GLRLM_HGRE	No Acute Xerostomia	20	14.40	0.295	42.000	12.80	0.421	46.000	14.25	0.387	45.000	
	Xerostomia	6	10.50			15.83			11.00			
GLRLM_SRLGE	No Acute Xerostomia	20	13.25	0.790	55.000	12.90	0.494	48.000	13.85	0.700	53.000	
	Xerostomia	6	14.33			15.50			12.33			
GLRLM_SRHGE	No Acute Xerostomia	20	14.25	0.387	45.000	13.65	0.882	57.000	13.85	0.700	53.000	
	Xerostomia	6	11.00			13.00			12.33			
GLRLM_LRLGE	No Acute Xerostomia	20	14.25	0.387	45.000	12.80	0.421	46.000	13.60	0.929	58.000	
	Xerostomia	6	11.00			15.83			13.17			
GLRLM_LRHGE	No Acute Xerostomia	20	14.65	0.176	37.000	14.70	0.157	36.000	14.15	0.457	47.000	

	Xerostomia	6	9.67			9.50			11.33			
GLRLM_GLNU	No Acute Xerostomia	20	13.15	0.700	53.000	13.45	0.976	59.000	13.45	0.976	59.000	
	Xerostomia	6	14.67			13.67			13.67			
GLRLM_RLNU	No Acute Xerostomia	20	13.40	0.929	58.000	13.35	0.882	57.000	13.40	0.929	58.000	
	Xerostomia	6	13.83			14.00			13.83			
GLRLM_RP	No Acute Xerostomia	20	13.35	0.880	57.000	13.10	0.656	52.000	13.65	0.882	57.000	
	Xerostomia	6	14.00			14.83			13.00			
NGLDM_Coarsenes s	No Acute Xerostomia	20	13.20	0.744	54.000	13.55	0.976	59.000	13.10	0.656	52.000	
	Xerostomia	6	14.50			13.33			14.83			
NGLDM_Contrast	No Acute Xerostomia	20	14.40	0.295	42.000	14.50	0.242	40.000	14.55	0.219	39.000	
	Xerostomia	6	10.50			10.17			10.00			
NGLDM_Busyness	No Acute Xerostomia	20	13.35	0.880	57.000	13.50	1.000	60.000	13.90	0.656	52.000	
	Xerostomia	6	14.00			13.50			12.17			
GLZLM_SIZE	No Acute Xerostomia	20	13.85	0.700	53.000	13.10	0.656	52.000	14.15	0.457	47.000	
	Xerostomia	6	12.33			14.83			11.33			

GLZLM_LZE	No Acute Xerostomia	20	13.20	0.740	54.000	13.30	0.836	56.000	13.40	0.929	58.000
	Xerostomia	6	14.50			14.17			13.83		
GLZLM_LGZE	No Acute Xerostomia	20	12.65	0.320	43.000	12.80	0.421	46.000	13.85	0.700	53.000
	Xerostomia	6	16.33			15.83			12.33		
GLZLM_HGZE	No Acute Xerostomia	20	14.20	0.421	46.000	13.40	0.929	58.000	13.95	0.614	51.000
	Xerostomia	6	11.17			13.83			12.00		
GLZLM_SZLGE	No Acute Xerostomia	20	13.30	0.836	56.000	12.50	0.242	40.000	13.45	0.976	59.000
	Xerostomia	6	14.17			16.83			13.67		
GLZLM_SZHGE	No Acute Xerostomia	20	14.20	0.421	46.000	13.75	0.790	55.000	13.75	0.790	55.000
	Xerostomia	6	11.17			12.67			12.67		
GLZLM_LZLGE	No Acute Xerostomia	20	13.10	0.656	52.000	13.30	0.836	56.000	13.45	0.976	59.000
	Xerostomia	6	14.83			14.17			13.67		
GLZLM_LZHGE	No Acute Xerostomia	20	13.10	0.656	52.000	13.30	0.836	56.000	13.70	0.836	56.000
	Xerostomia	6	14.83			14.17			12.83		
GLZLM_GLNU	No Acute Xerostomia	20	13.70	0.836	56.000	12.35	0.176	37.000	13.05	0.614	51.000
	Xerostomia	6	12.83			17.33			15.00		

GLZLM_ZLNU	No Acute Xerostomia	20	13.65	0.882	57.000	14.20	0.421	46.000	13.60	0.929	58.000
	Xerostomia	6	13.00			11.17			13.17		
GLZLM_ZP	No Acute Xerostomia	20	12.85	0.457	47.000	13.50	1.000	60.000	13.90	0.656	52.000
	Xerostomia	6	15.67			13.50			12.17		

Table 6.6: Comparison of patients with and without acute xerostomia based on the absolute paired percentage difference of all patients for each radiomic feature. Xerostomia was measured at mid-treatment. Radiomic features extracted from CT imaging taken before the start of treatment (CT1), mid-way through treatment (CT2), and at end of treatment (CT3). Statistical significance is based on independent Mann Whitney U test and values in bold indicate a significant result ($p < 0.05$).

Radiomic Feature	Event/Group (Xerostomia at Near End Treatment)	N	CT1 and CT2			CT2 and CT3			CT1 and CT3		
			Mean Rank	p value	U statistic	Mean Rank	p value	U statistic	Mean Rank	p value	U statistic
HISTO_Skewness	No Acute Xerostomia	23	13.57	0.940	33.000	14.43	0.095	13.000	13.78	0.648	28.000
	Acute Xerostomia	3	13.00			6.33			11.33		
HISTO_Kurtosis	No Acute Xerostomia	23	13.57	0.940	33.000	14.52	0.064	11.000	13.74	0.705	29.000
	Acute Xerostomia	3	13.00			5.67			11.67		
HISTO_ExcessKurtosis	No Acute Xerostomia	23	13.57	0.940	33.000	14.52	0.064	11.000	13.70	0.762	30.000
	Acute Xerostomia	3	13.00			5.67			12.00		
HISTO_Entropy_log10	No Acute Xerostomia	23	13.39	0.880	32.000	13.87	0.541	26.000	13.43	0.940	33.000
	Acute Xerostomia	3	14.33			10.67			14.00		
HISTO_Entropy_log2	No Acute Xerostomia	23	13.39	0.880	32.000	13.87	0.541	26.000	13.43	0.940	33.000
	Acute Xerostomia	3	14.33			10.67			14.00		
HISTO_Energy	No Acute Xerostomia	23	13.57	0.940	33.000	14.17	0.242	19.000	13.52	1.000	34.000

	Acute Xerostomia	3	13.00			8.33			13.33			
SHAPE_Volume_mL	No Acute Xerostomia	23	13.65	0.821	31.000	14.39	0.113	14.000	14.26	0.182	17.000	
	Acute Xerostomia	3	12.33			6.67			7.67			
SHAPE_Volume_vx	No Acute Xerostomia	23	13.65	0.821	31.000	14.39	0.113	14.000	14.26	0.182	17.000	
	Acute Xerostomia	3	12.33			6.67			7.67			
SHAPE_Sphericity	No Acute Xerostomia	23	14.74	0.018	6.000	13.87	0.541	26.000	13.74	0.705	29.000	
	Acute Xerostomia	3	4.00			10.67			11.67			
SHAPE_Surface	No Acute Xerostomia	23	13.52	1.000	34.000	13.48	1.000	34.000	14.35	0.134	15.000	
	Acute Xerostomia	3	13.33			13.67			7.00			
SHAPE_Compacity	No Acute Xerostomia	23	13.96	0.442	24.000	14.13	0.275	20.000	14.35	0.134	15.000	
	Acute Xerostomia	3	10.00			8.67			7.00			
GLCM_Homogeneity	No Acute Xerostomia	23	13.30	0.762	30.000	13.91	0.490	25.000	13.09	0.490	25.000	
	Acute Xerostomia	3	15.00			10.33			16.67			
GLCM_Energy	No Acute Xerostomia	23	13.48	1.000	34.000	14.00	0.395	23.000	13.65	0.821	31.000	
	Acute Xerostomia	3	13.67			9.67			12.33			
GLCM_Contrast	No Acute Xerostomia	23	13.39	0.880	32.000	13.61	0.880	32.000	12.96	0.352	22.000	

	Acute Xerostomia	3	14.33			12.67			17.67			
GLCM_Correlation	No Acute Xerostomia	23	12.70	0.157	16.000	12.52	0.078	12.000	12.65	0.134	15.000	
	Acute Xerostomia	3	19.67			21.00			20.00			
GLCM_Entropy_log10	No Acute Xerostomia	23	13.26	0.705	29.000	13.87	0.541	26.000	13.35	0.821	31.000	
	Acute Xerostomia	3	15.33			10.67			14.67			
GLCM_Entropy_log2	No Acute Xerostomia	23	13.26	0.705	29.000	13.87	0.541	26.000	13.35	0.821	31.000	
	Acute Xerostomia	3	15.33			10.67			14.67			
GLCM_Dissimilarity	No Acute Xerostomia	23	13.57	0.940	33.000	13.65	0.821	31.000	13.43	0.940	33.000	
	Acute Xerostomia	3	13.00			12.33			14.00			
GLRLM_SRE	No Acute Xerostomia	23	13.39	0.880	32.000	13.96	0.442	24.000	13.22	0.648	28.000	
	Acute Xerostomia	3	14.33			10.00			15.67			
GLRLM_LRE	No Acute Xerostomia	23	13.52	1.000	34.000	13.91	0.490	25.000	13.30	0.762	30.000	
	Acute Xerostomia	3	13.33			10.33			15.00			
GLRLM_LGRE	No Acute Xerostomia	23	13.83	0.594	27.000	14.09	0.312	21.000	13.70	0.762	30.000	
	Acute Xerostomia	3	11.00			9.00			12.00			
GLRLM_HGRE	No Acute Xerostomia	23	13.39	0.880	32.000	13.43	0.940	33.000	13.13	0.541	26.000	

	Acute Xerostomia	3	14.33			14.00			16.33			
GLRLM_SRLGE	No Acute Xerostomia	23	13.83	0.594	27.000	14.57	0.052	10.000	13.57	0.940	33.000	
	Acute Xerostomia	3	11.00			5.33			13.00			
GLRLM_SRHGE	No Acute Xerostomia	23	13.43	0.940	33.000	13.26	0.705	29.000	13.17	0.594	27.000	
	Acute Xerostomia	3	14.00			15.33			16.00			
GLRLM_LRLGE	No Acute Xerostomia	23	13.57	0.940	33.000	13.87	0.541	26.000	13.83	0.594	27.000	
	Acute Xerostomia	3	13.00			10.67			11.00			
GLRLM_LRHGE	No Acute Xerostomia	23	12.91	0.312	21.000	14.22	0.211	18.000	12.70	0.157	16.000	
	Acute Xerostomia	3	18.00			8.00			19.67			
GLRLM_GLNU	No Acute Xerostomia	23	13.39	0.880	32.000	14.26	0.182	17.000	14.09	0.312	21.000	
	Acute Xerostomia	3	14.33			7.67			9.00			
GLRLM_RLNU	No Acute Xerostomia	23	13.78	0.648	28.000	14.3	0.157	16.000	14.22	0.211	18.000	
	Acute Xerostomia	3	11.33			7.33			8.00			
GLRLM_RP	No Acute Xerostomia	23	13.48	1.000	34.000	13.91	0.490	25.000	13.26	0.705	29.000	
	Acute Xerostomia	3	13.67			10.33			15.33			
NGLDM_Coarseness	No Acute Xerostomia	23	13.91	0.490	25.000	13.57	0.940	33.000	14.22	0.211	18.000	

	Acute Xerostomia	3	10.33			13.00			8.00		
NGLDM_Contrast	No Acute Xerostomia	23	12.87	0.275	20.000	14.09	0.312	21.000	13.04	0.442	24.000
	Acute Xerostomia	3	18.33			9.00			17.00		
NGLDM_Busyness	No Acute Xerostomia	23	13.13	0.541	26.000	13.57	0.940	33.000	13.00	0.395	23.000
	Acute Xerostomia	3	16.33			13.00			17.33		
GLZLM_SZE	No Acute Xerostomia	23	13.96	0.442	24.000	14.48	0.078	12.000	14.17	0.242	19.000
	Acute Xerostomia	3	10.00			6.00			8.33		
GLZLM_LZE	No Acute Xerostomia	23	13.43	0.940	33.000	14.09	0.312	21.000	13.74	0.705	29.000
	Acute Xerostomia	3	14.00			9.00			11.67		
GLZLM_LGZE	No Acute Xerostomia	23	13.87	0.541	26.000	13.91	0.490	25.000	13.61	0.880	32.000
	Acute Xerostomia	3	10.67			10.33			12.67		
GLZLM_HGZE	No Acute Xerostomia	23	13.52	1.000	34.000	13.30	0.762	30.000	13.09	0.490	25.000
	Acute Xerostomia	3	13.33			15.00			16.67		
GLZLM_SZLGE	No Acute Xerostomia	23	13.78	0.648	28.000	13.96	0.442	24.000	13.48	1.000	34.000
	Acute Xerostomia	3	11.33			10.00			13.67		
GLZLM_SZHGE	No Acute Xerostomia	23	13.30	0.762	30.000	13.35	0.821	31.000	13.26	0.705	29.000

	Acute Xerostomia	3	15.00			14.67			15.33			
GLZLM_LZLGE	No Acute Xerostomia	23	13.70	0.762	30.000	14.09	0.312	21.000	13.74	0.705	29.000	
	Acute Xerostomia	3	12.00			9.00			11.67			
GLZLM_LZHGE	No Acute Xerostomia	23	13.48	1.000	34.000	14.09	0.312	21.000	13.87	0.541	26.000	
	Acute Xerostomia	3	13.67			9.00			10.67			
GLZLM_GLNU	No Acute Xerostomia	23	14.48	0.078	12.000	14.22	0.211	18.000	14.43	0.095	13.000	
	Acute Xerostomia	3	6.00			8.00			6.33			
GLZLM_ZLNU	No Acute Xerostomia	23	14.00	0.395	23.000	13.70	0.762	30.000	13.91	0.490	25.000	
	Acute Xerostomia	3	9.67			12.00			10.33			
GLZLM_ZP	No Acute Xerostomia	23	13.26	0.705	29.000	14.17	0.242	19.000	13.09	0.490	25.000	
	Acute Xerostomia	3	15.33			8.33			16.67			

Table 6.7: Comparison of patients with and without acute xerostomia based on the absolute paired percentage difference of all patients for each radiomic feature. Xerostomia was measured near the end of treatment. Radiomic features extracted from CT imaging taken before the start of treatment (CT1), mid-way through treatment (CT2), and at end of treatment (CT3). Statistical significance is based on independent Mann Whitney U test. Significant results ($p < 0.05$) are highlighted in bold.

Radiomic Feature	Event/Groups (Xerostomia 3-6 months post Treatment)	N	CT1 CT2			CT2 CT3			CT1 CT3		
			Mean Rank	p value	U statistic	Mean Rank	p value	U statistic	Mean Rank	p value	U statistic
HISTO_Skewness	No Late Xerostomia	17	12.12	0.220	53.000	12.24	0.263	55.000	13.65	0.916	74.000
	Late Xerostomia	9	16.11			15.89			13.22		
HISTO_Kurtosis	No Late Xerostomia	17	12.29	0.287	56.000	12.06	0.200	52.000	13.71	0.874	73.000
	Late Xerostomia	9	15.78			16.22			13.11		
HISTO_ExcessKurtosis	No Late Xerostomia	17	12.29	0.287	56.000	12.06	0.200	52.000	13.76	0.833	72.000
	Late Xerostomia	9	15.78			16.22			13.00		
HISTO_Entropy_log10	No Late Xerostomia	17	13.94	0.711	69.000	12.82	0.560	65.000	13.94	0.711	69.000
	Late Xerostomia	9	12.67			14.78			12.67		
HISTO_Entropy_log2	No Late Xerostomia	17	13.94	0.711	69.000	12.82	0.560	65.000	13.94	0.711	69.000
	Late Xerostomia	9	12.67			14.78			12.67		
HISTO_Energy	No Late Xerostomia	17	14.53	0.367	59.000	12.47	0.367	59.000	13.71	0.874	73.000
	Late Xerostomia	9	11.56			15.44			13.11		

SHAPE_Volume_mL	No Late Xerostomia	17	11.35	0.051	40.000	12.12	0.220	53.000	12.53	0.396	60.000
	Late Xerostomia	9	17.56			16.11			15.33		
SHAPE_Volume_vx	No Late Xerostomia	17	11.35	0.051	40.000	12.12	0.220	53.000	12.53	0.396	60.000
	Late Xerostomia	9	17.56			16.11			15.33		
SHAPE_Sphericity	No Late Xerostomia	17	10.71	0.009	29.000	12.94	0.634	67.000	13.53	1.000	76.000
	Late Xerostomia	9	18.78			14.56			13.44		
SHAPE_Surface	No Late Xerostomia	17	11.41	0.058	41.000	12.24	0.263	55.000	11.06	0.025	35.000
	Late Xerostomia	9	17.44			15.89			18.11		
SHAPE_Compacity	No Late Xerostomia	17	13.76	0.833	72.000	12.65	0.458	62.000	13.65	0.916	74.000
	Late Xerostomia	9	13.00			15.11			13.22		
GLCM_Homogeneity	No Late Xerostomia	17	12.00	0.181	51.000	11.53	0.075	43.000	12.76	0.525	64.000
	Late Xerostomia	9	16.33			17.22			14.89		
GLCM_Energy	No Late Xerostomia	17	13.94	0.711	69.000	12.06	0.200	52.000	12.65	0.458	62.000
	Late Xerostomia	9	12.67			16.22			15.11		
GLCM_Contrast	No Late Xerostomia	17	13.53	1.000	76.000	14.29	0.491	63.000	15.00	0.181	51.000
	Late Xerostomia	9	13.44			12.00			10.67		

GLCM_Correlation	No Late Xerostomia	17	13.65	0.916	74.000	14.59	0.339	58.000	13.59	0.958	75.000
	Late Xerostomia	9	13.22			11.44			13.33		
GLCM_Entropy_log10	No Late Xerostomia	17	13.18	0.792	71.000	12.47	0.367	59.000	13.76	0.833	72.000
	Late Xerostomia	9	14.11			15.44			13.00		
GLCM_Entropy_log2	No Late Xerostomia	17	13.18	0.792	71.000	12.47	0.367	59.000	13.76	0.833	72.000
	Late Xerostomia	9	14.11			15.44			13.00		
GLCM_Dissimilarity	No Late Xerostomia	17	12.94	0.634	67.000	14.35	0.458	62.000	14.47	0.396	60.000
	Late Xerostomia	9	14.56			11.89			11.67		
GLRLM_SRE	No Late Xerostomia	17	11.88	0.148	49.000	11.35	0.051	40.000	12.29	0.287	56.000
	Late Xerostomia	9	16.56			17.56			15.78		
GLRLM_LRE	No Late Xerostomia	17	12.12	0.220	53.000	11.12	0.029	36.000	11.88	0.148	49.000
	Late Xerostomia	9	16.11			18.00			16.56		
GLRLM_LGRE	No Late Xerostomia	17	12.00	0.181	51.000	12.76	0.525	64.000	13.88	0.751	70.000
	Late Xerostomia	9	16.33			14.89			12.78		
GLRLM_HGRE	No Late Xerostomia	17	14.18	0.560	65.000	13.35	0.916	74.000	14.29	0.491	63.000
	Late Xerostomia	9	12.22			13.78			12.00		

GLRLM_SRLGE	No Late Xerostomia	17	11.94	0.164	50.000	12.06	0.200	52.000	14.12	0.597	66.000
	Late Xerostomia	9	16.44			16.22			12.33		
GLRLM_SRHGE	No Late Xerostomia	17	13.94	0.711	69.000	14.53	0.367	59.000	14.00	0.672	68.000
	Late Xerostomia	9	12.67			11.56			12.56		
GLRLM_LRLGE	No Late Xerostomia	17	13.24	0.833	72.000	13.18	0.792	71.000	12.88	0.597	66.000
	Late Xerostomia	9	14.00			14.11			14.67		
GLRLM_LRHGE	No Late Xerostomia	17	14.35	0.458	62.000	12.29	0.287	56.000	14.29	0.491	63.000
	Late Xerostomia	9	11.89			15.78			12.00		
GLRLM_GLNU	No Late Xerostomia	17	12.47	0.367	59.000	12.71	0.491	63.000	13.00	0.672	68.000
	Late Xerostomia	9	15.44			15.00			14.44		
GLRLM_RLNU	No Late Xerostomia	17	11.53	0.075	43.000	12.59	0.426	61.000	12.82	0.560	65.000
	Late Xerostomia	9	17.22			15.22			14.78		
GLRLM_RP	No Late Xerostomia	17	12.00	0.181	51.000	11.00	0.021	34.000	12.06	0.200	52.000
	Late Xerostomia	9	16.33			18.22			16.22		
NGLDM_Coarseness	No Late Xerostomia	17	12.24	0.263	55.000	12.59	0.426	61.000	13.65	0.916	74.000
	Late Xerostomia	9	15.89			15.22			13.22		

NGLDM_Contrast	No Late Xerostomia	17	15.88	0.029	36.000	14.24	0.525	64.000	16.06	0.018	33.000
	Late Xerostomia	9	9.00			12.11			8.67		
NGLDM_Busyness	No Late Xerostomia	17	13.00	0.672	68.000	14.00	0.672	68.000	14.24	0.525	64.000
	Late Xerostomia	9	14.44			12.56			12.11		
GLZLM_SZE	No Late Xerostomia	17	10.82	0.013	31.000	11.71	0.107	46.000	12.65	0.458	62.000
	Late Xerostomia	9	18.56			16.89			15.11		
GLZLM_LZE	No Late Xerostomia	17	13.12	0.751	70.000	11.24	0.039	38.000	12.35	0.312	57.000
	Late Xerostomia	9	14.22			17.78			15.67		
GLZLM_LGZE	No Late Xerostomia	17	12.35	0.312	57.000	13.59	0.958	75.000	14.12	0.597	66.000
	Late Xerostomia	9	15.67			13.33			12.33		
GLZLM_HGZE	No Late Xerostomia	17	13.94	0.711	69.000	15.35	0.095	45.000	13.82	0.792	71.000
	Late Xerostomia	9	12.67			10.00			12.89		
GLZLM_SZLGE	No Late Xerostomia	17	12.06	0.200	52.000	13.12	0.751	70.000	14.71	0.287	56.000
	Late Xerostomia	9	16.22			14.22			11.22		
GLZLM_SZHGE	No Late Xerostomia	17	14.06	0.634	67.000	14.47	0.396	60.000	13.41	0.958	75.000
	Late Xerostomia	9	12.44			11.67			13.67		

GLZLM_LZLGE	No Late Xerostomia	17	12.47	0.367	59.000	11.47	0.066	42.000	12.71	0.491	63.000
	Late Xerostomia	9	15.44			17.33			15.00		
GLZLM_LZHGE	No Late Xerostomia	17	14.00	0.672	68.000	11.35	0.051	40.000	13.18	0.792	71.000
	Late Xerostomia	9	12.56			17.56			14.11		
GLZLM_GLNU	No Late Xerostomia	17	11.24	0.039	38.000	12.12	0.220	53.000	13.35	0.916	74.000
	Late Xerostomia	9	17.78			16.11			13.78		
GLZLM_ZLNU	No Late Xerostomia	17	11.65	0.095	45.000	12.29	0.287	56.000	12.65	0.458	62.000
	Late Xerostomia	9	17.00			15.78			15.11		
GLZLM_ZP	No Late Xerostomia	17	11.88	0.148	49.000	12.82	0.560	65.000	13.06	0.711	69.000
	Late Xerostomia	9	16.56			14.78			14.33		

Table 6.8: Comparison of patients with and without late xerostomia based on the absolute paired percentage difference of all patients for each radiomic feature. Xerostomia was measured 3-6 months following the completion of treatment. Radiomic features extracted from CT imaging taken before the start of treatment (CT1), mid-way through treatment (CT2), and at end of treatment (CT3). Statistical significance is based on independent Mann Whitney U test and values in bold indicate a significant result ($p < 0.05$).

The absolute paired percentage difference of GLCM_Correlation observed in the first half of treatment (CT1 and CT2) was significantly different for patients who experienced acute xerostomia measured at mid-treatment ($U = 25.000, p = 0.033$). However, the differences observed for this specific feature value between CT1 and CT2 was not associated with xerostomia measured at the near end of treatment ($U = 16.999, p = 0.157$) or 3-6 month following the completion of treatment ($U = 74.000, p = 0.916$). The absolute paired percentage difference observed in the first half of treatment (CT1 and CT2) for Shape_Sphericity was associated with acute xerostomia reported at the 'near end of treatment' ($U = 6.000, p = 0.018$).

Examining the variation of radiomic feature values over the whole course of treatment (i.e. CT1 and CT3), two features were associated with the development of late xerostomia; they are SHAPE_Surface ($U = 35.000, p = 0.025$) and NGLDM_Contrast ($U = 33.000, p = 0.018$), with the latter also showing an association with late xerostomia when feature variability was measured in the first half of treatment ($U = 36.000, p = 0.029$).

In addition, the variation of three other radiomic feature values observed during the first half of treatment (CT1 and CT2) were also associated with late xerostomia, but interestingly, not during the second half. These were SHAPE_Sphericity ($U = 29.000, p = 0.009$), GLZLM_SZE ($U = 31.000, p = 0.013$) and GLZLM_GLNU ($U = 38.000, p = 0.039$).

The variation of three radiomic feature values observed during the second half of treatment (CT2 and CT3) were associated with late xerostomia measured 3-6 months following the completion of treatment, none of which showed an association with late xerostomia based on the observations during either the first half of treatment (CT1 and CT2), or the treatment course overall (CT1 and CT3). These were GLRLM_LRE ($U = 36.000, p = 0.029$), GLRLM_RP ($U = 34.000, p = 0.021$,) and, GLZLM_LZE ($U = 38.000, p = 0.039$).

The Mann Whitney U test results presented in Tables 6.6-6.8 confirms the hypothesis that for some individual radiomic features there is a statistically

significant difference in the absolute paired percentage difference in feature values for the patients who experienced xerostomia, compared to those who did not. The significant features which have been isolated are displayed in Figures 6.3 – 6.6.

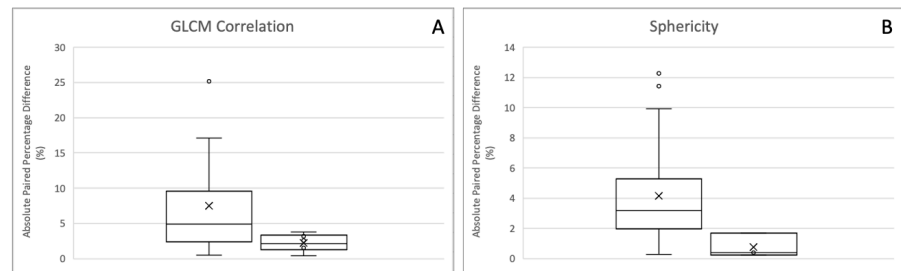


Figure 6.3: Box and Whisker plots of features in which there was a statistically significant difference in the absolute paired percentage difference from CT1 to CT2 for the patients who experienced A1 xerostomia (Fig. 3a), or A2 xerostomia (Fig. 3b) compared to those who did not. Plots are arranged to display “No Xerostomia” on the left and “Xerostomia on the right”.

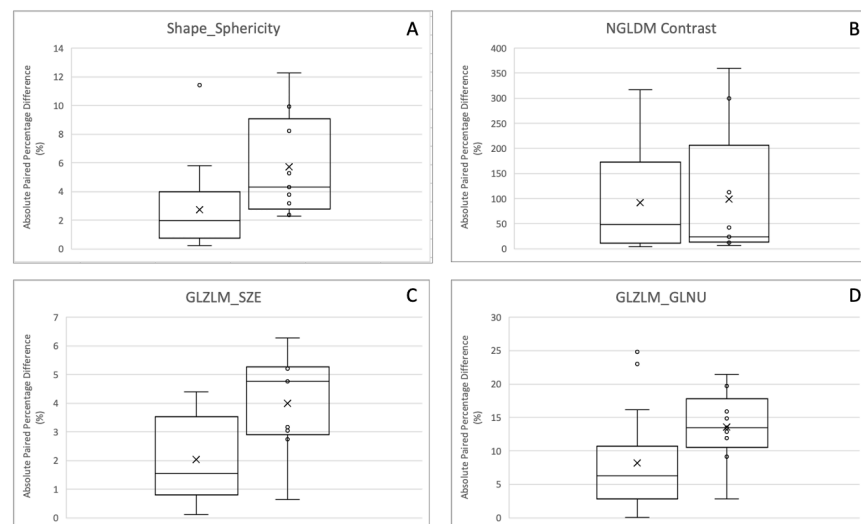


Figure 6.4: Box and Whisker plots of features in which there was a statistically significant difference in the absolute paired percentage difference from CT1 to CT2 for the patients who experienced late

xerostomia compared to those who did not. Plots are arranged to display “No Xerostomia” on the left and “Xerostomia on the right”.

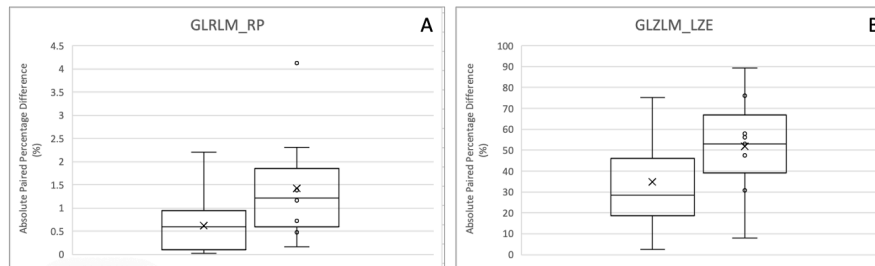


Figure 6.5: Box and Whisker plots of features in which there was a statistically significant difference in the absolute paired percentage difference from CT2 to CT3 for the patients who experienced late xerostomia compared to those who did not. Plots are arranged to display “No Xerostomia” on the left and “Xerostomia on the right”

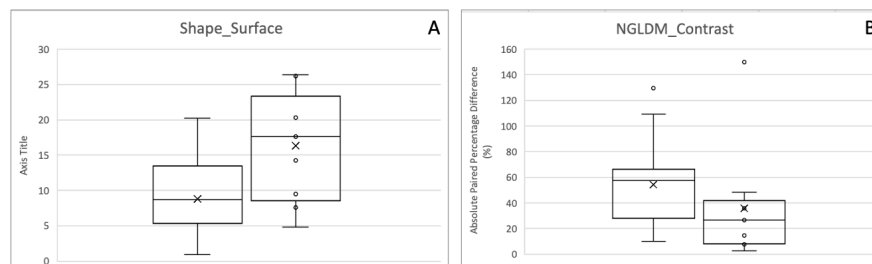


Figure 6.6: Box and Whisker plots of features in which there was a statistically significant difference in the absolute paired percentage difference from CT1 to CT3 for the patients who experienced late xerostomia compared to those who did not. Plots are arranged to display “No Xerostomia” on the left and “Xerostomia on the right”. NB: Two patients from the "No Xerostomia" group were excluded from the plot as they were outliers, with an absolute paired percentage difference in 654% and 1127%.

6.3.4 The Impact Of The Size, Shape, Sphericity, and Compacity of the Parotid Glands on Xerostomia

Figures 6.7 – 6.10 display the paired percentage difference for the Shape family features per patient.

6.3.4.1 *Shape_Volume (mL)*

Figure 6.7a indicates that in the first half of treatment, the average size of the parotid glands (measure in mL) decreased in 21 patients (81%) with a mean paired percentage difference for the whole cohort of -10%. In the second half of treatment (Fig. 6.7b) there was an additional decrease in parotid gland volume in 19 patients (73%) with a mean paired percentage difference for the whole cohort of -7%.

Of the six patients who reported acute xerostomia at mid treatment (A1), five had a reduction in the paired percentage difference of the average volume of their parotid glands in the first half of treatment (CT1 and CT2), and all six had a reduction over the full course of treatment (CT1 and CT3). The average volume of the parotid glands decreased in two out of the three patients who reported acute xerostomia near the end of treatment (A2) during both the first and second halves of treatment. All three patients with A2 xerostomia experienced an overall reduction in their average parotid gland volume over the full course of treatment (CT1 and CT3). With respect to late xerostomia, of the nine patients who reported this toxicity, nine and six of them reported a decrease in parotid gland volume in the first and second halves of treatment respectively, with eight reporting a reduction over the full course of treatment (CT1 and CT3).

Looking at specific patients, Patient 1 (P1) experienced the most parotid gland shrinkage, with a decrease of 23% and 32% in the first and second half of treatment respectively. This patient reported experiencing xerostomia during the middle of their treatment, and 3-6 months following,

however not at the completion of their treatment. Two patients (P15 and P17) concluded their treatment with an overall increase in their parotid gland volume of 2% and 7% respectively between CT1 and CT3. Neither of these patients reported experiencing acute xerostomia at either time point (mid or near-end treatment), but one reported late xerostomia measured 3-6 months following the completion of treatment.

6.3.4.2 Shape_Surface

Similar results can be observed for the Shape_Surface feature which reflects the surface area of the region of interest (Fig 6.8). In the first half of treatment (Fig 6.8a), there was a reduction in the surface area of the average parotid glands in 20 patients (77%), reflected in the negative paired percentage difference between CT1 and CT2, which for the whole cohort was -7%. In the second half of treatment, the surface area of the average parotids decreased in 21 patients (81%); however, this value was <1% for four patients (15%) and the average paired percentage difference for all patients was -3%.

Of the six patients who reported acute xerostomia at mid treatment (A1), four had a reduction of the average surface area of their parotid glands in the first half of treatment (CT1 and CT2), and five had a reduction over the full course of treatment (CT1 and CT3). Of the three patients who reported acute xerostomia near the end of treatment (A2), two experienced a reduction in the average surface area of the parotid glands during both the first and second halves of treatment. All three patients with A2 xerostomia experienced an overall reduction in their average parotid gland surface area over the full course of treatment (CT1 and CT3). With respect to late xerostomia, of the nine patients who reported this toxicity, nine and seven of them reported a decrease in parotid gland surface area in the first and second halves of treatment respectively. In agreement with results observed for average parotid volume (mL), eight of the nine patients with

late xerostomia experienced a reduction in average parotid surface area over the full course of treatment (CT1 and CT3).

As depicted in Figure 6.8c, over the full course of treatment the paired percentage difference was negative in all but three patients (P15, P17, and P25). Of these three patients, regarding the incidence of xerostomia, no clear pattern can be realised; with P25 reporting xerostomia at just the mid-treatment time point (A1), P17 reporting just late xerostomia with no incidence of acute toxicity, and P15 not reporting any xerostomia at all during or following treatment.

6.3.4.3 Shape_Sphericity

The Shape_Sphericity feature describes how spherical an ROI is, and the results for this feature are displayed in Figure 6.9. In the first half of treatment the paired percentage difference in the Shape_Sphericity feature was less than 10% in 23 patients (88%) with an average for the cohort of 0.4%. Of the three patients with a paired percentage difference which varied greater than 10% from CT1 to CT2, two reported a decrease in the glands' average sphericity, and one reported an increase. Of the two patients with this early decrease in sphericity, both (P1 and P3) reported late xerostomia, and one (P1) reported acute xerostomia at the mid-treatment time point. In the second half of treatment (CT2 and CT3), one patient's Shape_Sphericity value showed a greater than 10% deviation, with an 18% increase; however, this same patient (P3) had a 12% decrease in the average sphericity of their parotid glands during the first half of treatment, and an overall 4% increase when comparing variation between the first and last CT scans (CT1 and CT3). This patient reported the presence of late xerostomia, without a mention of any acute xerostomia during treatment, possibly due to poor documentation during treatment rather than a lack of acute toxicity experienced by the patient.

Four out of the six patients with acute xerostomia mid-way through treatment (A1) reported a decrease in the average sphericity of their

parotid glands during both the first half of treatment, and also over the full course of treatment. Acute xerostomia experienced at the near end of treatment (A2) was reported by three patients, with changes in the average sphericity noted in one of these patients over the full course of treatment (CT1 and CT3) as well as the second half of treatment specifically (CT2 and CT3). For the nine patients who reported late xerostomia, the reduction in the parotid glands average sphericity was noted in two patients during the first half of treatment; however, this figure rose to seven patients in the second half of treatment.

6.3.4.4 Shape_Compacity

Shape_Compacity reflects how compact the ROI is. There was a loss of average compacity of parotid glands in 16 patients (62%) during the first half of treatment (CT1 and CT2), and 22 patients (85%) during the second half of treatment (CT2 and CT3) as shown in Figure 6.10. In total, 13 (50%) and 16 (62%) patients had a paired percentage difference within 10% during the first half and second half of treatment respectively. The average paired percentage difference of Shape_Compacity for the whole cohort was -3% and -4% for the first and second halves of treatment respectively.

All six patients with early acute xerostomia (A1), reported a decrease in this feature value during both the first half of treatment (CT1 and CT2) and over the full course of treatment (CT1 and CT3), reflected by the negative paired percentage difference. Two out of the three patients who experienced acute xerostomia near the end of treatment (A2), experienced a reduction in the average compacity of their parotid glands both during the first and second half of treatment. Five out of nine patients with late xerostomia experienced a reduction in glands compacity during the first half of treatment (CT1 and CT2), and eight out of the nine patients also experienced this reduction during the second half of treatment (CT2 and CT3).

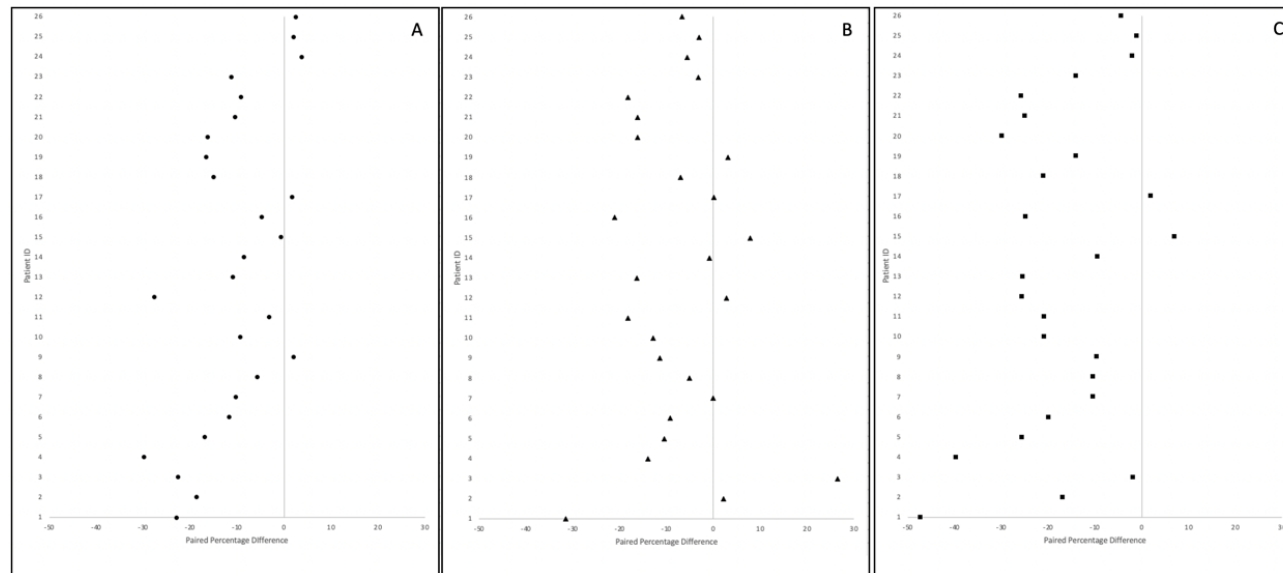


Figure 6.7: Paired percentage difference of **Shape_Volume (mL)** per patient measured from CT imaging taken before the start of treatment and at mid-way during treatment (A), mid-way through treatment and at the completion of treatment (B), and before the start of treatment and at the completion of treatment (C). A stable result, i.e., no change in feature value is indicated by the vertical line (0% paired percentage difference), data points plotted to the left of this vertical line indicate a reduction in the paired percentage difference whilst a data point plotted to the right of the vertical line indicates an increase in the paired percentage difference of the radiomic feature value per patient.

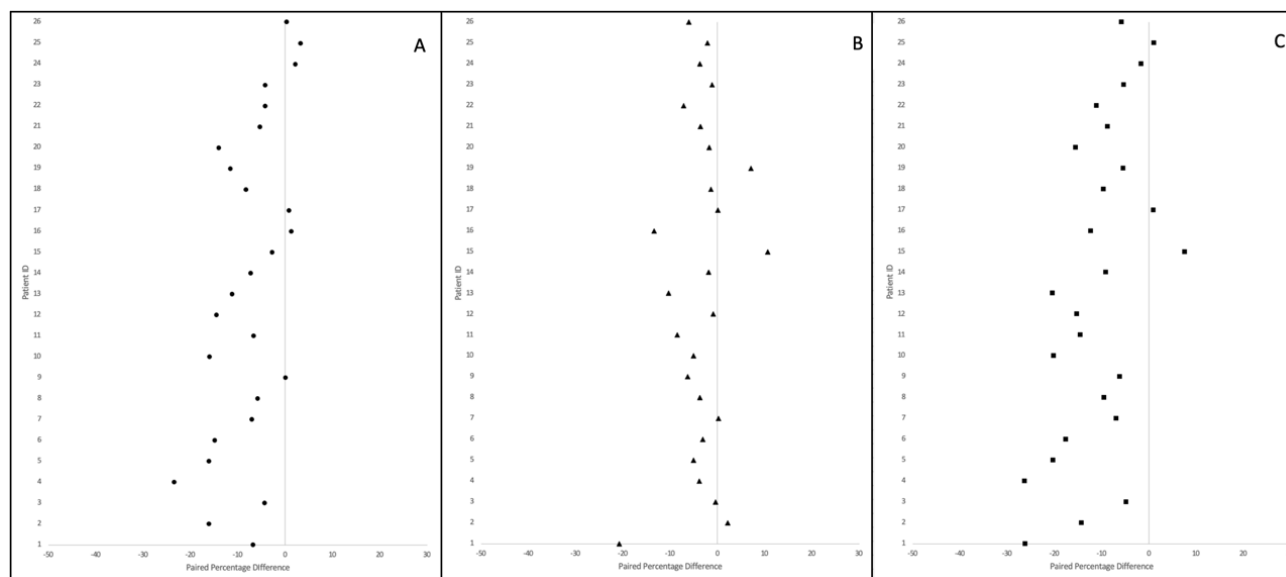


Figure 6.8: Paired percentage difference of **Shape_Surface** per patient measured from CT imaging taken before the start of treatment and at mid-way during treatment (A), mid-way through treatment and at the completion of treatment (B), and before the start of treatment and at the completion of treatment (C). A stable result, i.e., no change in feature value is indicated by the vertical line (0% paired percentage difference), data points plotted to the left of this vertical line indicate a reduction in the paired percentage difference whilst a data point plotted to the right of the vertical line indicates an increase in the paired percentage difference of the radiomic feature value per patient.

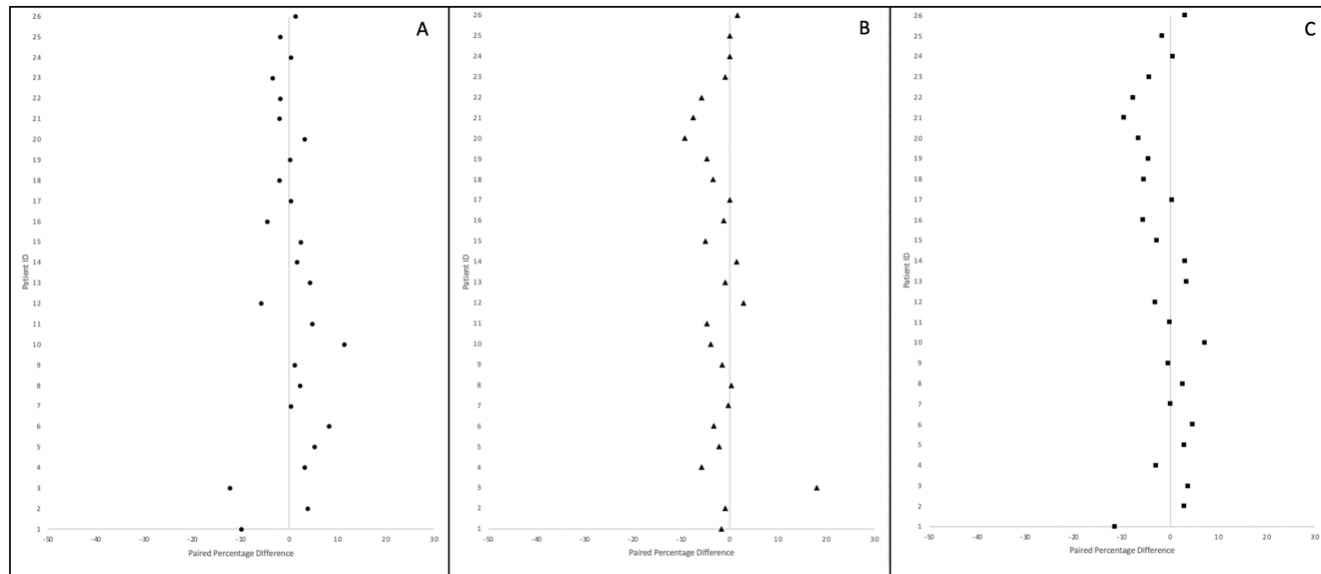


Figure 6.9: Paired percentage difference of **Shape_Sphericity** per patient measured from CT imaging taken before the start of treatment and at mid-way during treatment (A), mid-way through treatment and at the completion of treatment (B), and before the start of treatment and at the completion of treatment (C). A stable result, i.e., no change in feature value is indicated by the vertical line (0% paired percentage difference), data points plotted to the left of this vertical line indicate a reduction in the paired percentage difference whilst a data point plotted to the right of the vertical line indicates an increase in the paired percentage difference of the radiomic feature value per patient.

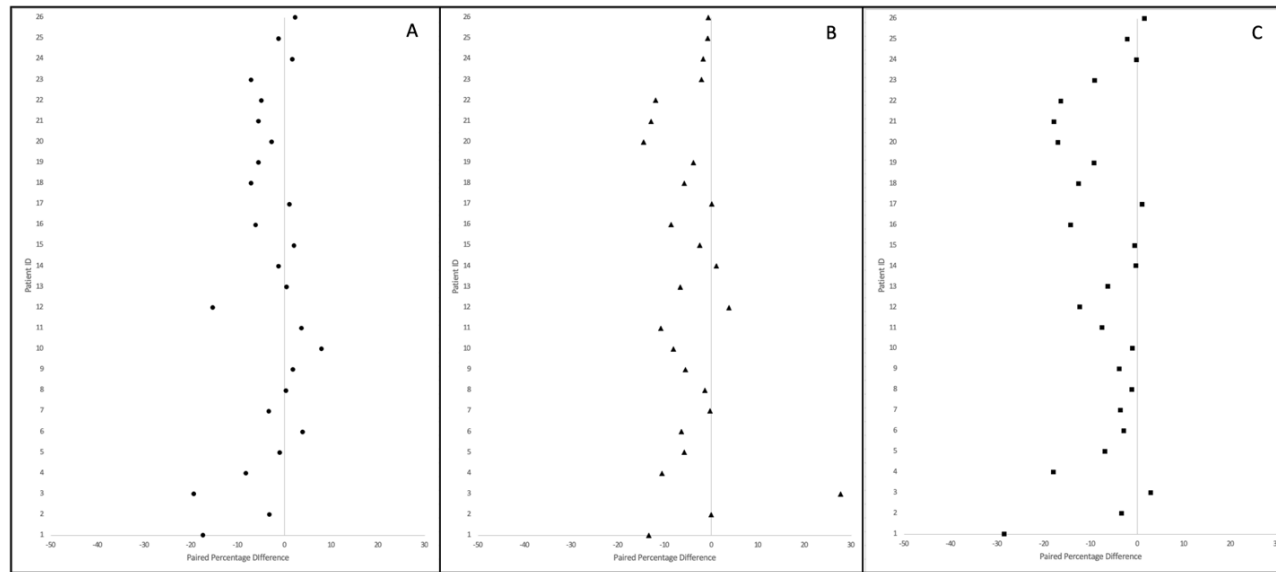


Figure 6.10: Paired percentage difference of **Shape_Compacity** per patient measured from CT imaging taken before the start of treatment and at mid-way during treatment (A), mid-way through treatment and at the completion of treatment (B), and before the start of treatment and at the completion of treatment (C). A stable result, i.e., no change in feature value is indicated by the vertical line (0% paired percentage difference), data points plotted to the left of this vertical line indicate a reduction in the paired percentage difference whilst a data point plotted to the right of the vertical line indicates an increase in the paired percentage difference of the radiomic feature value per patient.

6.4 DISCUSSION

This work has demonstrated the dynamic nature of parotid gland biology and morphology as a result of radiation; changes which were reflected in the variation of CT based radiomic features during a course of radiotherapy to the head and neck. While limited to a very small patient cohort, there appears to be statistically significant differences in radiomics features at time points when both acute and late xerostomia are reported.

6.4.1 Variation in Parotid Gland Morphology

The results presented in Figures 6.7 and 6.8 indicate that over the course of radiotherapy, on average the size of the parotid gland decreases, both in total volume and also surface area.

Based on more frequent imaging, several other studies have reported a reduction in parotid gland volume of approximately 30% during the course of treatment; noting that the ipsilateral gland experiences the greatest volume loss (232-235). Despite similar cohort sizes to these previous studies, the results presented here were more modest in which just two patients experienced an overall reduction of their average parotid gland volumes in excess of 30%, and the mean reduction in parotid gland volume of this cohort was just 17%. Loo et al., (233) reported a mean percentage reduction of 30.2% for the ipsilateral parotid over the full course of treatment. Unfortunately, in their study the entirety of the parotid glands was not imaged due to the limited length of the MVCT on Tomotherapy, and the authors concede that the portion of the parotid gland superior to the external auditory canal was not included in the field of view in any of the patients studied. Furthermore, although they reported a mean percentage reduction of 30.2% for the ipsilateral parotid, with a range of 17.1% to 41.5% in a cohort of just 5 patients, the potential influence of outliers in their results cannot be ignored. Nevertheless, this percentage

volume reduction is consistent with Bhide et al., (232) who used more frequent imaging on a larger patient cohort (n=20). Given the relatively small size of the parotid glands, Bhide et al., (232) also reported the glands' reduction in absolute volume (cm³) to complement the percentage reduction results. In the first two weeks of treatment, the greatest reduction in gland volume was 4.2cm³ with a subsequent reduction of 4.0cm³ between weeks two and four. These results are more closely aligned with what was observed here. In absolute terms, the greatest reduction in the average of the glands' volumes in the first half of treatment was 7.4mL, and 6.6mL in the second half of treatment. Unfortunately, all of the other studies discussed reported the reduction based on percentage alone and did not include absolute volumes, making direct comparison with results presented here difficult (234, 235). Furthermore, although Bhide et al., (232) do report the absolute reduction in parotid gland volume, they do not provide the volume of the glands at baseline. In fact, none of the studies discussed provide baseline details of the parotids' volume. A final key difference in the work here, is the fact that the ipsilateral and contralateral parotid glands were not analysed separately, but rather the mean value of the two parotid glands per patient were reported. This approach was taken to be consistent with previous similar radiomic studies (144).

Important to note is that none of the above studies mentioned (232-235) used radiomic analysis to examine the impact of radiotherapy on parotid gland *shape* but rather simply described the size and shift in the glands centre of mass (position) during a course of treatment. Radiomics affords us the opportunity to extract additional quantitative detail regarding the morphology and underlying texture of the parotid glands. By examining first order shape features, not just the volume of the gland, but also the surface area, sphericity, and compacity can be quantified. In this study, unsurprisingly, similar results were observed for the Shape_Surface as for the Shape_Volume features. Conversely, sphericity of the glands is not influenced to the same magnitude as the volume and surface area. The mean sphericity of the patient cohort was stable throughout the course of

292

treatment with an average increase of just 0.4%. The glands' compacity showed a decrease, with 50% (n=13) and 62% (n=16) of patients having a paired percentage difference within 10% during the first half and second half of treatment respectively. These results indicate structural changes within the gland during treatment reflective of acinar cell loss in response to radiation damage, coupled with an increase in adipose tissue within the ROI (144, 236). The low incidence of xerostomia reported in this cohort may therefore be rationalised by the modest reduction in the glands' compacity and potentially inconsequential loss of acinar cells experienced in this patient group.

Using an "in-house" approach with Python, Scalco et al., also investigated the value added from more comprehensive quantitative features to describe structural change in the parotid gland. In their study, during the course of radiotherapy they observed a 3.5% ($p = 0.0002$) decrease in local entropy estimated from the co-occurrence matrix, indicating a less complex tissue organization in response to radiation. In the work presented here, variation in the GLCM_Entropy feature was also observed based on both the Friedman test, and also Wilcoxon signed rank tests considering the individual timepoints. Similar to the earlier work of Fiorino et al., (236), Scalco et al., (144) did not aim to correlate their findings with a clinical endpoint, and despite the agreed variation in parotid GLCM_Entropy of their study with data presented here, this specific radiomic feature was not associated with either acute or late toxicity based on these data. That is not to say that parotid tissue organisation can be disregarded with respect to xerostomia, but rather points to possible limitations of this work (i.e., low patient numbers, and low incidence of events and subjective nature of the endpoint, xerostomia). Furthermore, there are specific clinical factors that need to be considered, for example, the impact of bilateral nodal irradiation on toxicity as well as the specific chemotherapy agents used. The patient cohort used in this study was mixed with respect to disease site, clinical stage as well as cytotoxic drugs used, all of which would impact on the results.

6.4.2 Impact of Variation of Radiomic Features on Xerostomia

A meta-analysis published in 2018 reported that despite the use of modulated techniques, the incidence of acute and late xerostomia was still as high as 80% and 75% respectively (87). In contrast to this, the overall incidence of acute and late toxicity of this patient group was low with incidence rates of just 23% and 35% for acute (A1) and late xerostomia respectively; possibly owing to the modest reduction in parotid gland volume and surface area comparative to other studies. Historically, predicting the incidence of xerostomia has been limited to dose-volume parameters which are used in normal tissue complication probability (NTCP) models, and only recently has there been an interest in considering the underlying biology, which is possible with radiomic analysis.

There are a limited number of studies that have identified radiomic features measured at baseline to be predictive of xerostomia (116, 129, 185). Whilst novel in their use of radiomics to predict xerostomia, these studies did not consider the dynamic nature of the parotid glands in response to radiation (237), and as such disregard the deformation of the gland that was established in earlier work.

Studies using longitudinal imaging and “delta-radiomics” serve to bridge this gap. Pota et al., (177) was one of the first studies to confirm CT radiomic features to be predictive of xerostomia 12 months following radiotherapy in a cohort of 37 patients. Similar to this study, they extracted radiomic features based on CT imaging taken at three time points (before treatment, at the middle of treatment, and at the end of treatment). They concluded that the final volume variation measured over a course of treatment was the best single feature for classifying severity of late xerostomia; however, they did not specify parotid volume at baseline therefore limiting direct comparison with results here in Study 4. Teshima et al., (238) previously established that a reduction in parotid volume is correlated with a reduction in salivary flow. Conversely, based on the results presented here, the change in Shape_Volume (mL) feature

measured over the full course of treatment was not associated with late xerostomia ($U = 60.000$, $p = 0.396$); however, this provides an opportunity to consider the subtle differences in endpoints measured. Teshima et al., (238) quantitatively measured salivary flow based on the Saxon test, and yet outcomes used in this study used the more subjective sensation of xerostomia. Despite the differences in parotid volume, there is greater agreement regarding the other Shape features. The variation in the surface area (SHAPE_Surface) was associated with late xerostomia ($U = 35.000$, $p = 0.025$) in this study, and these findings are supported by the work of van Dijk et al., (239) who identified the change in the contralateral parotid gland surface in a group of 68 patients to be the most predictive feature of xerostomia measured 12 months following the completion of treatment. Again, it must be noted that despite the agreement in specifying the importance of surface variation, van Dijk et al., (239) isolated the contralateral gland explicitly, whereas the work presented here averaged out the radiomic feature values between the ipsi- and contralateral glands; a decision guided by the results of the Wilcoxon Signed Rank test results in Study 3 of this thesis. Regardless, van Dijk et al., (239) highlighted that it was the change in parotid gland surface measured at week 3 specifically (i.e. mid treatment), that had the strongest association with late xerostomia ($p < 0.001$). Regarding these early treatment changes, the current study showed that it was the variation in the gland's sphericity in the first half of treatment that was most strongly associated with late xerostomia ($U = 29.000$, $p = 0.009$). This result indicates that the major structural "damage" to the parotid glands which occurs early during treatment can have significant consequential late xerostomia. This is aligned with the work by Gabrys et al., (185) in which a multivariable analysis of clinical, dosiomic and, radiomic features also identified the gland's eccentricity (i.e. sphericity) to be strongly associated with the incidence of late xerostomia; however, their analysis was based on features extracted at baseline, and did not include a delta analysis.

Similar to the work here, Van Dijk et al., (239) also considered acute xerostomia measured at week 3; however, there was not a significant

association between the change in parotid gland surface during the first half of treatment and acute xerostomia ($p = 0.14$). Likewise, data presented here did not demonstrate an association between the variation of Shape_Surface measured from CT1 and CT2, and mid-treatment acute (A1) xerostomia ($U = 59.000, p = 0.976$).

In the same year that van Dijk et al., (239) published their delta-radiomics work, so too did Liu et al., (178) in which they turned their attentions to acute toxicities alone. Using CT data of 35 patients with nasopharyngeal cancer they acquired images every 10th fraction during treatment, as well as 3 months following treatment. In addition to the commonly used outcome of xerostomia based on the RTOG acute scoring system, unique to their study was the addition of the more objective endpoint of salivary output; both measured in 10-day intervals. They identified eight features which were able to predict salivary output measured at fraction 10, and 14 features which predicted salivary output measured at fraction 30. This was based on an insignificant difference ($p > 0.05$) between the predicted salivary output and the measured output. Relevant to the current work is the delta-radiomics component (i.e., the change in feature values from baseline to fraction 10) that contributed significantly to the prediction of salivary output. It should be noted however, that the radiomic features in the study by Liu et al., (178) were all extracted for the left and right parotid glands separately. Whilst van Dijk et al., (239) also considered the glands separately, they only reported results for the contralateral gland as these features performed better in their predictive models than the ipsilateral gland.

As with the current results, Pota et al., (177) and van Dijk et al., (239) both identified at least one feature from the shape “family” to be associated with xerostomia. This association between xerostomia and shape features is in agreement with historical observations of studies conducted in the pre-radiomics era and is logical considering the dosimetric consequence a variation in gland size can have during the course of treatment. However, results from Study 1 indicated that the Shape_Volume feature in particular

is influenced by inter-observer delineation variability of the parotid gland, and more robust textural features could be more reliable (240).

It has already been established that several texture features measured at baseline have the capacity to predict xerostomia (116, 129); however, there is less agreement in the literature regarding the predictive power of texture features when measured with longitudinal imaging. In the delta-radiomics study by van Dijk et al., (239) the authors clearly state that of the intensity and texture features selected, none were significantly associated with late xerostomia regardless of what time point the change in feature value was measured. Conversely, of the features that Liu et al., (178) isolated as being predictive of salivary output based on feature variability over the course of treatment, none were intensity or volume based features. The authors venture the reason for the apparent preference of higher order features was due to the large ratio that wavelets had in their analysis, which included over 1700 radiomic features. However, inherent differences in the patient groups examined may also provide an explanation. For example, patient age is an important factor to consider, whereby the density of the gland decreases with increasing age; likely due to an increase in adipose tissue (241). The median age of the 35 patients studied by Liu et al., (178) was 48 (range: 24 to 72), and whilst van Dijk (239) do not report individual patient ages, 30% of their patients were over the age of 65. Comparatively, this figure is close to the median age of patients presented in this study (66; range 42 to 87). More recently Berger et al., (242) demonstrated improved prediction model performance for xerostomia with the inclusion of texture analysis based on MVCT imaging of 117 patients. Unique to their study was the use of daily imaging analysis allowing for a more complete view of radiomic “kinetics” in comparison to the majority of studies, including this one, that performed a longitudinal analysis based on limited timepoints.

Whilst there are fundamental methodological differences between this study and others (177, 178, 239), what is a common observation to all is that variability of radiomic features within the first half of treatment is associated with both acute and late xerostomia. In broad terms, whilst

there may be commonalities observed with respect to feature families, there continues to be a lack of agreement regarding the individual radiomic features. In addition to the more obvious methodological differences, other possible explanations for this lack of agreement may be related to the patients included in the individual studies. These differences may be intrinsic to the patient; for example age, as previously discussed, or the presence of existing co-morbidities such as Sjogren's syndrome, hypothyroidism, anxiety or depression, all of which are known to be associated with xerostomia (243). Tumour site and stage of disease influence volumes irradiated, and hence risk of toxicity. Heterogeneity of these clinical features present in this patient cohort undoubtedly had an impact on results observed in this study. Alternatively, extrinsic factors may be at play; for example the use of certain medications which are known to have a detrimental impact on salivary flow and would influence not only baseline xerostomia levels (244), but also affect the structure of the patient's parotid glands; however, without access to the patients' medical records, these details are unknown here and the impact on results is purely speculative.

6.4.3 Limitations of Study 4

There are limitations of this work which must be acknowledged. The numbers in the patient cohort are limited but are similar to those reported in other studies. However, as there are very few published studies with longitudinal imaging of the parotid gland over the treatment schedule, this is a unique strength of this analysis.

A second limitation of this work is the retrospective nature and reliance of data already collected and annotated within the patient file. In the clinical notes available to us from the Cancer Imaging Archive, there was no recording of xerostomia at baseline, prior to the commencement of treatment. This is potentially problematic considering a systematic review published in 2010 reported that between 6-12% of patients presented with xerostomia at baseline (226). This lack of detail in the clinical notes could

have also occurred during the course of treatment, with variations in the level of detail noted from one clinician to another is likely. There may have also been some missing data, and where xerostomia was not recorded in the notes, it cannot be assumed that the patient was not actually experiencing this toxicity at the time. Although it is recognised that toxicity reporting is notoriously poor with patients not enrolled on a clinical trial, the classifying missing data as G0 toxicity is an inappropriate assumption. Furthermore, it is likely patients who experienced moderate to severe acute xerostomia would have been appropriately managed, minimising consequential late effects resulting in an absence of reportable toxicity.

Finally, the scale used to assess xerostomia in the patient group must also be addressed. Although the RTOG Acute Toxicity Scale is a validated and widely used tool, it is well recognised that xerostomia is a subjective sensation and the use of patient reported outcomes measures may be more appropriate here. Although PROMs are also susceptible to missing data they are potentially a more informative and meaningful method of capturing the impact radiotherapy induced xerostomia has on the quality of life of our patients (245).

6.5 CONCLUSION

This study has described the changes in radiomic features of the parotid gland during a course of radiotherapy, and has identified features which are associated with acute and late xerostomia based on CT imaging. This work has once again demonstrated the added value of incorporating radiomics for the biological analysis of the parotid glands' response to radiation that extends beyond simple volumetric changes. In doing so, this study has highlighted the potential for the incorporation of radiomics to support clinical decisions regarding adaptive radiotherapy.

7.1 ADVANCING RADIATION TOXICITY PREDICTION THROUGH RADIOMICS

Through the integration of radiomic analysis of routine CT imaging, this thesis has addressed the limited availability of tools which provide a more personalised approach for the prediction of radiation induced parotid gland damage during and following radiotherapy to the head and neck. The central hypothesis of this thesis was that the quantification of morphological and textural composition of the parotid gland, made possible with radiomic analysis of CT images, can predict acute radiation induced xerostomia and sticky saliva. Whilst the potential of radiomics to serve as a decision support tool is becoming more widely recognised, the process is computationally complicated, requiring robust methodology prior to integration into a clinical workflow (246). As such, the central hypothesis of this thesis was tested across four specific studies which addressed radiomic feature variability under different conditions, as well as tested the predictive power of radiomics, and finally explored the dynamic nature of radiomic feature values in response to radiation.

Study 1: Influence of Inter-observer Delineation Variability on Radiomic Features of the Parotid Gland

Prior to image pre-processing and radiomic feature extraction, the region of interest (ROI) is defined. Region of interest, or tumour and organ delineation, is made easier with the addition of multimodality imaging and contouring guidelines; however, despite these aids, the influence of the observer themselves (translated as inter-observer delineation variability) remains. Studies in radiotherapy have quantified the impact of inter-

observer delineation variability on dosimetric outcomes (247), yet far fewer have examined this critical step in the context of radiomics. The impact of inter-observer delineation variability on radiomic features was therefore poorly understood and identified a critical knowledge gap in the literature. Furthermore, given some studies indicated the downstream impact of delineation variability on the predictive performance of radiomic based models (163), this issue needed to be examined prior to moving forward with this thesis. The work conducted in the first study of this thesis served to bridge this knowledge gap.

Within this thesis, Study 1 tested the hypothesis that some radiomics features are more robust against inter-observer delineation variability than others, rendering them more appropriate for clinical use. Using thresholds of feature robustness previously identified in the literature, Study 1 ultimately identified a subset of 14 robust features, the majority of which were derived from the Grey Level Run Length Matrix (GLRLM) family. Features within the GLRLM family provide information regarding the size of homogenous runs for each grey level, essentially quantifying the heterogeneity of the parotid gland tissue, or in other words the irregularity of the parotid parenchyma (129). The identification of these features in this dataset is supported by previous studies which also reported individual features within this family to improve the performance of predictive models when identifying patients more likely to experience chronic xerostomia (193). The GLRLM_SRE feature describes the distribution of the short homogenous runs within a ROI and GLRLM_RLNU describes the length of the homogenous runs in a ROI. These two features were regarded clinically relevant for the prediction of xerostomia by van Dijk et al., (184) and Nardone et al., (129) respectively. In the context of predicting xerostomia, the fact that these two features were also found to be robust to observer delineation variability based on the results of Study 1 further amplifies their likely clinical relevance. These features could be used to reliably reflect the composition of the gland irrespective of the anatomical boundaries defined by an observer.

Aside from supporting earlier reports the results of Study 1 informed the exploration of their clinical link within this thesis (Studies 3 and 4). A good example is Shape_Sphericity. In Study 1 this feature was robust to inter-observer variability with a mean paired percentage difference of $4.75\% \pm 0.03\%$. A modest increase in this feature during the first half of treatment was later associated in Study 4, with both acute and late xerostomia; establishing its robustness to delineation variability along with its clinical value. The differences in the bin width used between Study 1 and Study 4 which could have impacted on the feature value, were unlikely to affect these results, given Shape_Sphericity was found to be reproducible irrespective of bin width in Study 2.

Study 1 was the first to consider the impact of ROI delineation on radiomic feature values in a normal tissue and the first to bring these considerations to the parotid glands. Until then, the literature was dominated by publications in which the lung gross tumour volume was the ROI (157, 158, 163, 171, 248).

The results of this work must be considered within the context of early studies which aimed at assessing radiation induced changes to the parotid glands, without consideration for potential variability caused by observer delineation variability (116, 210). In a modelling study using CT and MRI radiomic features to predict xerostomia and sticky saliva, Sheikh et al., (116) defended this methodological shortcoming by arguing the importance of developing modules which reflect common clinical practice; i.e. multiple observers in the clinic. Considering this argument, it must be acknowledged that “common clinical practice” has evolved to include the routine use of auto-segmentation, and as such the use of auto-segmentation was also integrated into Study 1. The results here demonstrated that overall, there was a higher percentage of robust features when auto-segmentation methods were used, rather than manual delineation (10% vs 7%). This conclusion, also supported by previous studies (158), informed the methodology for Study 2 in which the parotid contours were generated using the Smart Segmentation workspace in the Varian Eclipse treatment

planning system. The integration of auto-segmentation into radiomics studies has been subsequently embraced by others in a range of cancer sites and imaging modalities (249-251).

Beyond improvements observed in feature stability with auto-contouring, other gains can be seen in this approach. Radiomic studies typically requires large patient cohorts to enhance statistical power and generalisability of results, and manually contouring these cases can be extremely time consuming. By making use of auto-contouring solutions, observer delineation variability is reduced, as is the time taken to contour the ROIs. Ferrante et al., (250) and Dutta et al., (251) both utilised contours generated by deep learning methods showing equivalent results with clinician generated contours, thereby accelerating their radiomics workflow.

The work of Study 1 was based on radiomic data extracted from a single patient, with a single scan from a single institution. Image pre-processing was limited in this study, in which the native voxel size was preserved with no spatial resampling, and the default bin width of fixed bin 10 HU was used. This approach, whilst not flawed for a single patient study did raise the question how radiomic feature values might change based on pre-processing settings, such as voxel size and bin width. These questions were subsequently addressed in Study 2 of the thesis.

Study 2: Influence of Spatial Resampling Combined with Intensity Discretization on Reproducibility of Radiomic Features of the Parotid Gland

Translating radiomics into a clinically useful tool has been hampered by technical challenges partly related to a lack of standardisation in feature extraction and a lack of harmonisation in imaging acquisition and reconstruction protocols (246). To address these issues, Mali et al., (200) broadly groups harmonisation strategies into either feature domain or image domain. Harmonising in the feature domain refers to identifying

reproducible features via test-retest methods, or alternatively using statistical approaches to compensate for differences in feature values because of differing image acquisition protocols. The latter may include simple approaches such as z-score normalisation or more sophisticated statistical methods such as combatting batch effects (ComBat) (252). This approach takes its origin from genomics studies; however, in the context of radiomic studies, it attempts to correct the uncontrollable errors that can occur as a result of systematic technical differences in the imaging acquisition and reconstruction protocols (253). Alternatively, researchers can attempt to harmonise these uncertainties prior to feature extraction, i.e., harmonising in the “image domain” as suggested by Mali et al., (200). This process involves post-processing techniques, such as image resampling and intensity histogram binning. Whilst these specific techniques were both identified as standard operating procedures in the 2023 review by Huang et al., (246), the impact of these harmonisation techniques on radiomic feature values was poorly understood in earlier research.

Study 2 addressed the ongoing challenge of harmonising imaging protocols when working with multi-institutional imaging data. Spatial resampling to a common voxel size is recommended by the IBSI to align patients’ scans to the same frame of reference, as well as enabling accurate texture analysis given many features assume symmetry in voxel size (125). The need for spatial resampling may have been acknowledged at the time Study 2 was conducted; however, was not common practice in radiomic studies which examine the parotid glands as the ROI. The results from Study 2, therefore, addressed this methodological “blind spot”. From the results of this study, it was found that the majority of radiomic features (60%) were influenced by spatial resampling. The necessity of this step is exacerbated when using multi-institutional data where image reconstruction methods and image resolution varies, highlighting the need to include spatial resampling into the standard operating procedures for radiomic studies (135, 246).

Whilst the need for spatial resampling was perhaps better understood, the influence of image intensity discretization and, in particular, the selection

of bin width was not clearly appreciated when parotid glands are the ROIs. Some authors simply relied on existing head and neck radiomic literature to inform these parameters (116); however, methodologies varied for example, imaging type and ROIs examined. As a result, the validity of accepting these parameters as protocol is questioned. From the results of the secondary aim in Study 2, it was seen that the majority of radiomic features (58%) were not influenced by grey-level discretization; however, given this study only examined two bin sizes of similar size this result may have varied if more, larger bin widths were explored. Whilst the bin widths used in Study 2 were smaller than that of other parotid based radiomic studies, selecting a bin width which is too large risks features being heavily averaged out resulting in data loss (116, 135).

In 2020 van Timmeren et al., (135) conceded that bin width selection remained an open question. The results presented in Study 2 provided some clarity on this topic. When investigating the combined impact of intensity discretization with image resampling, over a third of radiomic features saw an improvement in their “robustness classification” when the bin width was reduced from 10 HU to 5 HU. Feature robustness is critical in radiomics research to ensure results are generalisable across different imaging platforms and patient groups, thereby reaffirming their clinical application. Robustness of individual features have been shown to be influenced by several factors related to image acquisition, reconstruction, ROI delineation and image pre-processing; factors Reiazi et al., (136) identify as “disruptive parameters”. The results presented in both Study 1 and Study 2 of this thesis identify features robust to observer delineation variability and reconstruction parameters. Although Study 1 utilised a fixed bin width of 10 HU exclusively, five common features which were robust to the “disruptive parameters” investigated in each study were identified. Namely, these include Shape_Sphericity, GLCM_Inverse Difference, GLRLM_SRE, GLRLM_RP, and GLSZM_SZE. Of these, we are reminded of the clinical relevance of GLRLM_SRE when predicting xerostomia, previously identified by van Dijk et al., (184).

Whilst this study was informative for advancing the understanding of spatial resampling and image discretization in an ROI previously under-researched, this reproducibility analysis lacked clinical endpoints and as such the relevance of these results was limited to hypothesis generating; limitations that were subsequently addressed in Studies 3 and 4.

Study 3: Prediction of Acute Xerostomia and Sticky Saliva Using Baseline CT Radiomic Analysis

Study 3 study aimed to improve the prediction of acute xerostomia and sticky saliva (SS) through the inclusion of radiomic analysis of the parotid glands for patients with head and neck cancer. When relying on clinical and dosimetric variables only, the predictive power of the logistic regression models generated in Study 3 for both xerostomia and sticky saliva were not much better than random chance. This confirms that the current clinical approach to predictive modelling is insufficient, and more novel methods to predict treatment outcomes and support clinical decisions must be investigated. At the time of Study 3, the use of radiomics to predict xerostomia, and to a lesser extent, sticky saliva had been explored by other researchers (116, 129, 184, 193). Collectively, they all demonstrated improvements in prediction of these side effects and the added value of integrating radiomics into their respective models. The radiomic work completed in Study 3 served to contribute to this field, which was rapidly expanding, by reaffirming the potential of radiomics for the prediction of xerostomia and sticky saliva.

When radiomic features were incorporated into the clinical model, the classification performance of these regression models improved for both xerostomia and sticky saliva as the outcomes. This clearly demonstrated the added value of radiomics and, despite using different radiomics software to others, confirmed what had been reported in the literature. Nardone et al., (129) Sheikh et al., (116) and van Dijk et al., (184) all investigated the potential predictive gain with the inclusion of radiomic features as imaging biomarkers for xerostomia, with the latter also

306

attempting to predict sticky saliva. In each study, the highest predictive power was observed for the models which included clinical, dosimetric, and radiomic features. Naturally, these promising results then raised the question if radiomics alone would be powerful enough to identify which patients are likely to experience more severe toxicity. The results here were less impressive compared to the combined models, but still demonstrated an improvement from the clinical only models. Realistically, the only gain from utilising a radiomics only model is related to missing data within patients' medical records that can occur in busy clinical environments. CT scans, and the data contained within them, are always captured as part of the radiotherapy planning process, as a result these data are never missing.

One critical step in the radiomics workflow that had been overlooked in previous predictive studies was the impact intensity discretization can have on these clinical outcomes. Linking back to the work conducted in Study 2, Study 3 demonstrated the predictive value of radiomics for xerostomia and sticky saliva, irrespective of bin width. These results addressed this previously ignored concept, hopefully prompting researchers new to the field to consider this detail in their own radiomics methodology. Whilst selecting a specific bin width is still a required step prior to feature extraction, the dependency of the predictive power of the models generated on the bin width is evidently less critical, thereby allowing researchers greater freedom in bin width selection (254). Despite this, it is still important for future researchers to declare the image discretization method as well as the resultant bin width or number of bins, and is a requirement for reporting according to the CheckList for the EvaluAtion of Radiomic Research (CLEAR) (217).

Using radiomics, Study 3 exploited data contained within CT images acquired at baseline for treatment planning purposes. Patients receiving chemoradiation to the head and neck region often experience anatomical changes in response to treatment, for example weight loss or tumour regression. Parotid shrinking during radiotherapy is also extremely common, and in the case of Ogura et al., (255) was observed in as many of

63% of parotid glands in a cohort of 41 patients. The reduction in the volume of the parotid gland is common and also not insignificant. Wu et al., (256) reported a mean reduction in parotid gland volume of 35% ($p < 0.001$), and although both Ogura et al., (255) and Wu et al., (256) used conventional radiotherapy, similar observations have been reported with the use of modulated techniques (257-259). As one example, Fiorentino et al., (259) quantified a mean daily loss in parotid volume of 1.5%, resulting in a total volume loss of up to 20% by the third week of a course of IMRT. Given the frequency and magnitude of parotid gland shrinkage during treatment, rescanning approximately midway through treatment for offline dosimetric analysis is not uncommon in clinical practice (259, 260). With these additional images acquired, it affords us the opportunity to examine these scans more closely to explore not just how the gross anatomy changes in response to radiation, but also radiomic features themselves; observations which were investigated in Study 4 of this thesis.

Study 4: Radiomic Analysis of Parotid Gland Transformations During Radiotherapy and the Association with Xerostomia

In the case of parotid glands, the radiation induced tissue damage that occurs is reflected in a shrinkage of the gland itself as a result of loss of gland parenchyma and atrophy of acinar cells (256). Whilst the change in total volume of the gland may be easily appreciated on CT imaging, the more complex vascular changes that occur are not so apparent. With the knowledge that saliva is produced by blood plasma (261) Wu et al., (256) overcome this challenge by utilising doppler ultrasound to assess the haemodynamic nature of the parotid glands following treatment. Their results demonstrated significant differences in vascular velocity, restrictive index and pulsarity index between radiotherapy patients and the control group. Whilst informative, the use of doppler ultrasound during a standard course of radiotherapy is not common, and given the data generated via radiomics we are afforded the opportunity to apply radiomic analysis of CT images to assess these complex changes.

The challenge of quantifying these changes on CT imaging has been addressed in Study 4 with the use of radiomic analysis. This study confirmed the expectation that radiomic features vary significantly over the course of treatment; an observation that was apparent in 95% of the features examined. When looking at the first and second half of treatment separately, it was apparent that the greatest variation in texture feature values occurs early during treatment. This result is somewhat expected considering patients can experience a 50-60% loss of salivary flow within the first week of treatment (61, 262). This is as a result of acinar cell loss which can occur within the first 10 days of radiotherapy (263, 264), and therefore not surprisingly is reflected in textural patterns quantified with radiomics.

The clinical relevance of the dynamic nature of the radiomic features became apparent considering the variability of several features were associated with either acute or late xerostomia. From a clinical perspective, mid-treatment is a common timepoint in which offline adaptive radiotherapy is considered, due to the anatomical response observed on imaging (265). Variability in features at this point which are associated with toxicity therefore provides us with a timely opportunity to adapt the plan. As presented in Study 4 of this thesis, variation in the Shape_Sphericity feature during the first two weeks of treatment is associated with xerostomia measured at the end of treatment.

This feature was also found to be robust to observer delineation variability in Study 1. Identifying this link between studies is important considering Study 4 examined images from “real world” data and likely suffered from delineation variability. In Study 4, the parotid glands were contoured based on local practice, which may have included some element of auto-segmentation under the supervision of several radiation oncologists, rather than a single observer. Although the voxel size and bin widths varied between Study 1 and Study 4, this feature was also robust to resampling irrespective of bin width based on results of Study 3. In addition, early variation in this feature along with variation of NGLDM_Contrast,

GLZLM_SIZE, and GLZLM_GLNU were all associated with xerostomia which persisted 3-6 months following treatment.

Interestingly, variability in none of these features were linked with acute xerostomia measured at the mid treatment time point. This implies that clinically relevant textural changes of the parotid gland can be observed via radiomic analysis prior to the onset of these late clinical symptoms. Using this information allows us to act early where new plans can be produced which include more conservative dose volume constraints for the parotid, although caution must be exercised to ensure adequate target coverage is maintained and the risk of local recurrence is negated (266). Regardless, this form of radiomic driven adaptive radiotherapy makes use of the full amount of data hidden within medical images taken in routine practice. The judgment to adapt a plan or not, is not always easy or clear. With patient selection criteria being so vast and varied (267), the decision often being physician dependant or made in an ad hoc manner. With the more objective data that radiomics affords us, these clinical decisions can be made more efficiently and with greater confidence.

7.2 THE POTENTIAL OF RADIOMICS IN ADAPTIVE RADIOTHERAPY

To decrease dosimetric uncertainties and improve clinical outcomes, radiotherapy often includes plan adaptation during a course of treatment. Adaptive radiotherapy involves modifying the plan during a course of treatment based on patient specific observations that were not considered during initial planning (268). Taking a proactive approach to offline ART, Castelli et al., (269) reported an 11% decrease in xerostomia risk with weekly replanning, and it is these clinical gains that have supported its widespread implementation. Depending on the approach taken, ART can be resource intensive; however, considering changes in head and neck

anatomy is gradual, typically evolving over a period of weeks, the less intensive “offline” approach to plan adaptation has been previously deemed appropriate for this site (270). The results presented in Study 4 indicate that these recommendations should be reconsidered in light of the rapid textural changes which occur prior to the clinical manifestation of toxicity.

The clinical practice of ART historically has focused primarily on anatomical variations that impact on the position of the target and surrounding structures. These observed variations may include weight loss (232), parotid gland shrinkage (271) or a reduction in the target volume (272). Brown et al., (273) included these anatomical features in their predictive model study, where patients with oropharyngeal cancer who were at a high risk for the need for ART included patients >100kg at the start of treatment, stage N3 disease or those with an initial node size >55mm. Whilst some of these patient and nodal features may seem extreme, they are important to consider in light of the changing patient profile observed over the years. With global obesity incidence nearly tripling between 1975 and 2016, we are seeing more and more patients with a weight exceeding 100Kg in our departments (274). Furthermore, based on Surveillance Epidemiology and End Results (SEER) data, several authors have reported a decrease in increase in node positive disease (275, 276). These trends reflect the clinical characteristics of HPV associated tumours which are not only becoming more common, but are also more likely to present with smaller primary tumours and more extensive nodal disease (8). In a systematic review of the literature, Brouwer et al., (267) contributed to the clinical characteristics identified by Brown et al., (273) by further adding age, body mass index, planned dose to the parotid glands, the initial parotid gland volume and the volume of the parotid gland overlapping the planning target volume.

This consideration of the position of the parotid gland and dose being received is vital in ensuring organ function is preserved and toxicity is minimised. The dosimetric impact of this parotid gland “migration” has

been demonstrated in several studies, all reporting an increase in either maximum parotid dose, mean parotid dose or parotid volume receiving 26 Gy or more (234, 277, 278). Based on clinical characteristics, several authors have attempted to identify early those patients who would be most likely to benefit from ART, suggesting replanning around week 3 or 4 of treatment would be the most appropriate (234, 273). Whilst this provides us with valuable information based on treatment imaging, the characteristics presented by Brown et al., (273) and Brouwer et al., (267) are purely clinical, and the decisions to adapt a plan are based purely on gross anatomical changes observed on imaging. There may be scope to utilise this imaging data in a more quantitative fashion, and it is here that we must consider the role of radiomics in both anatomical and biologically driven adaptive radiotherapy (279).

Several groups have integrated radiomics as a clinical decision tool, informing the need to trigger ART for head and neck patients (280-282). A 2023 review of the literature revealed radiomics has been used for both anatomical based ART, as well as biological (i.e. response) based ART; however, the focus still very much remains on the target and identifying early “regressors” (279). Focusing on the tumour volume exclusively rather than organs at risk is arguably short-sighted and does not reflect the patterns of practice of ART (POP-ART) based on a survey of 177 institutions (283). Results from the POP-ART survey revealed head and neck was the most common cancer site for ART, and reasons to adapt were based on a combination of target and organs at risk; highlighting the close interaction of these structures within the region observed in previous studies (284). For example, shrinkage of the tumour in head and neck radiotherapy often results in the parotid glands shifting medially into regions of high dose (267, 271, 284). As a result, replanning in response to this observation is likely to have a positive impact on the severity of xerostomia and sticky saliva (269).

Given the findings presented in this thesis, along with the current published evidence, predicting toxicity based on either the value of specific baseline radiomic features, or early changes reflected in radiomics feature

instability, could present a more holistic approach to ART which is triggered not only based on the tumour but also incorporates normal tissue changes. Furthermore, by incorporating radiomics as a decision support tool for ART allows for objectivity. Despite the quantitative metrics identified by Brouwer et al., (267) and Brown et al., (273) the decision to adapt treatment or not remains largely ad-hoc, and this subjectivity may be dampened with radiomic analysis (283, 285). Adapting patients' original treatment plans using an OAR focused and radiomics driven method will thereby decrease treatment burden on normal tissue, irrespective of observed tumour response to treatment. This use of advanced imaging, and this approach to personalised radiotherapy would have a direct impact on the side effects experienced by patients and hence impact on their quality of life, both during and following radiotherapy.

A radiomics driven approach to offline adaptive radiotherapy would not be too dissimilar to existing offline approaches described by Green et al., (286). The key difference however is that assessing the need for a replan would not be solely based on a dosimetric assessment but rather would incorporate extraction of radiomic features (Figure 7.1).

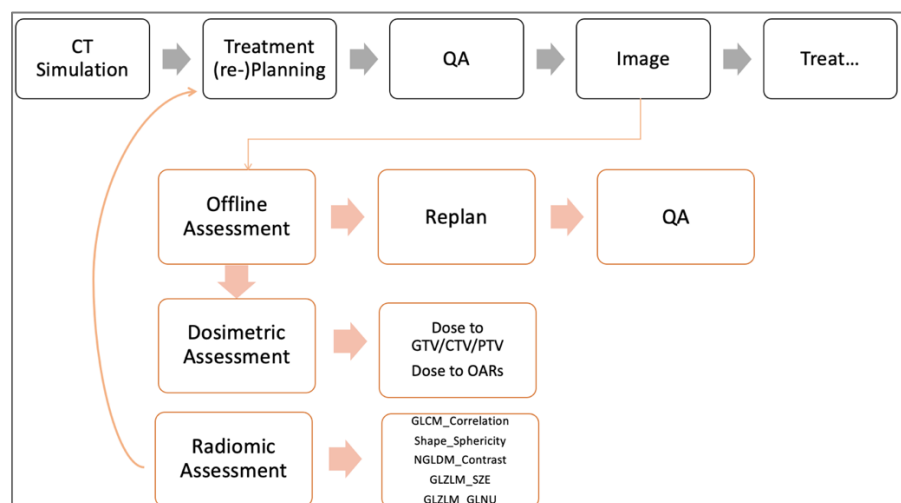


Figure 7.1: Radiomic driven offline adaptive workflow relevant radiomic features specified. These features were identified in Study 4 in which mean

paired percentage difference between baseline CT and mid-treatment CT were associated with either acute or late xerostomia.

Previous, non-radiomic studies have recommended replanning midway through treatment is sufficient to have a positive dosimetric and clinical impact on parotid gland sparing. Despite these recommendations, there may be a hesitancy to adopt these recommendations, especially if gross anatomy has remained stable by this point in time (287). The results of Study 4 support the narrative of mid-treatment replanning by demonstrating that the early variation in several radiomic features was associated with acute and late xerostomia. This recommendation is supported by van Dijk et al., (239) who performed a delta-radiomic analysis of the parotid glands based on weekly CT, concluding that textural changes noted mid-treatment significantly improves the prediction of late xerostomia.

7.3 INTEGRATING IMAGING BIOMARKERS WITHIN A CLINICAL AND SOCIAL CONTEXT

Unveiling hidden data through radiomic analysis exposes the complexity of the underlying anatomy and provides an opportunity for a deeper and more personalised approach to predicting treatment outcomes. The work in this thesis revealed some overarching trends, proved the central hypothesis presented, and identified some specific radiomic features of interest. Study 1 identified 14 features which were robust to observer delineation variability. Of these, four were also reproducible across different bin widths: Shape_Sphericity, GLCM_Inverse Difference, GLRLM_SRE, and GLRLM_RP. With respect to clinical relevance, the integration of radiomics improved the prediction of both xerostomia and sticky saliva irrespective of bin width. Of the nine features derived from the principal component analysis, five were also common across bin widths investigated. The dynamic behaviour of eight features were associated with either acute or late xerostomia. Of all features the GLRLM_RP appeared to be most reproducible and clinically relevant. The GLRLM feature family describes the size of homogenous runs for each grey level. A low run percentage as reflected as a low GLRLM_RP value, indicates a highly uniform or homogenous region of interest (125). Given the parotid gland is largely made up of adipose tissue containing acinar cells (263), variation in this feature value would reflect the changing composition of the parotid gland in response to radiation. The findings presented in this thesis, however, are of little value without considering them in the wider clinical and societal context.

7.3.1 The Individual and Social Burden of Xerostomia and Sticky Saliva

Despite technological and biological advances in radiotherapy, treatment related toxicities still present lifelong complications for patients. These

treatment related toxicities not only have a consequence for the individual patient's quality of life, but also for their families, carers, health care providers, and the wider society (288).

From the patient's perspective, xerostomia and sticky saliva present many challenges that impact on their activities of daily living and quality of life. More specifically, the consequences of these effects include mucositis, infection (such as candidiasis), speech and mastication difficulties, dysphagia, dental demineralisation, changes of taste, and nutritional deficiencies (289). Ge et al., (288) quantified the impact of radiotherapy-related toxicity on quality of life in patients with head and neck cancers via a meta-analysis; however, taking a more qualitative approach is equally valuable, especially considering xerostomia and sticky saliva are subjective conditions.

Jiang et al., (290) conducted semi-structured interviews with 20 patients to explore the impact of xerostomia from a more holistic perspective. In doing so, five domains emerged from the shared experiences and perspectives from patients. Communication problems were highlighted by these patients, prompting some to simply talk as little as possible. If left unmanaged, these new behaviours could easily develop into feeling of social isolation and depression. Psychosocial problems such as these also featured heavily in these interviews, with patients "feeling trapped by feelings of deep sadness, fear, anxiety and anguish, which were mainly caused by the persistent and uncontrolled dryness" (290). A multitude of physical problems were also identified, for example, loss of appetite and weight loss; both of which manifest from an inability to swallow and eat. Patients who experience severe radiation induced xerostomia often alter their diet not just during the acute phase but develop life-long behaviours that may contribute to increased risk of dental decay. These behaviours were observed in the 2024 study by Kawashima et al., (42), in which patients sought intraoral comfort by choosing more cariogenic food and beverages. Evidently, sweet taste recovers early post radiotherapy and so these new, "less healthy" lifestyle habits can develop quickly following

treatment (42, 291). For that reason, using radiomics to identify patients at greater risk of developing xerostomia could assist dietitians in working more closely with patients in need of educational intervention and increased dietary support. Although the systematic review by Ziedler et al., (292) was unable to make a definitive statement regarding the effectiveness of nutritional intervention for patients with head and neck cancer, five out of the six studies reviewed all reported improvement in patients' quality of life with increased support from a dietitian.

A subcategory identified from the qualitative analysis from Jiang et al., (290) pertained to "excessive costs", which sheds light on other hidden treatment related "toxicities". Toxicity from cancer treatment is multi-dimensional and includes not just clinical events, but also the monetary burden associated with side effect management. "Financial toxicity" refers to the direct and indirect medical expenses patients incur during their treatment and survivorship periods (293). The most obvious direct expense includes the actual cost of treatment which is not insignificant given head and neck cancer is often treated with tri-modality therapy; however, the financial burden of treatment related side effects is also considered a direct cost. These costs may include accessing ancillary medical services, or the purchasing of additional pharmaceuticals required for long term side effect management (294). One such example comes from the purchasing of saliva stimulants, which have proven to be beneficial for the management of oral health and xerostomia following cancer treatment (295). Furthermore, given the important role dentists have in preventing caries and periodontal disease as a result of hyposalivation, the costs associated with increased dental visits is another example of a direct cost of treatment (294, 296). Xerostomia and sticky saliva are known to impact on patients' ability to speak, and affected patients find themselves requiring specialised care from speech and language therapists; another direct out of pocket expense (294). Finally, mucositis is also a common side effect from chemoradiation to the head and neck, occurring in as many as 70% of patients and is exacerbated by decreased salivary flow (297). Extended hospital stays due to severe mucositis has also been identified as a significant direct cost from

head and neck cancer treatment (293, 297). Whilst this may be largely attributed to the chemotherapy component of patients' treatment, given chemotherapy is typically administered concurrently with radiation for head and neck cancers we cannot ignore the synergistic, financial impact of combined therapies which was highlighted in the univariate analysis conducted by D'Rummo et al., (298) .

Rosi-Schumacher et al., (293) identified loss of lifetime wages as one of the biggest contributors to indirect financial toxicity for patients treated for head and neck cancer. Most obviously, this indirect financial burden is due to patients missing work during treatment; however, the financial impact of treatment can continue following treatment completion. In this regard, the impact of xerostomia and sticky saliva might not seem clear, yet these toxicities were specifically highlighted by Verdonck-de Leeuw et al., (299) as a significant barrier to employment.

The value of radiomics to improve treatment related toxicity has been demonstrated by the work presented in this thesis, along with the literature (210). The impact of radiomics, however, with respect to alleviating *financial* toxicity is poorly understood. Huang et al., (246) briefly mentions the potential of radiomics to assist in directing clinical management in a way that financial toxicity is reduced; however, detailed evidence of this in the literature base is lacking. Social workers have been identified as key stakeholders in the care of cancer patients, with an established impact on easing financial toxicity (300). Considering this, Longo et al., (301) call for a proactive approach to be taken when referring patients to a social worker. Unfortunately, referral rates can be low and the reasons for this is multifactorial, including poor identification of patient specific issues (302). The objectivity that radiomics affords us may to some extent address this uncertainty by identifying patients at a higher risk of xerostomia and sticky saliva; side effects which impact on financial toxicity. Whilst integrating social support services into the routine management of all patients is recommended it is not always possible. When departments are under resourced and services are stretched, radiomics could supplement current

screening for financial toxicity and help guide management, thereby elevating it to a clinically useful tool as suggested by Huang et al. (246).

The financial toxicity as a result of these direct and indirect costs will, of course, vary depending on the country of treatment, yet regardless of whether these costs come out of the patients' pocket or the exchequer, they are not insignificant. Taking a more fiscal view, Elting and Chang (303) estimated the financial burden of oral complications from head and neck cancer treatment. Although the authors were unable to aggregate results through meta-analysis, costs reported in all studies were converted to a common currency using estimates computed based on the International Monetary Fund. Whilst the authors acknowledged the ongoing expenses of managing late xerostomia and sticky saliva could be lessened through the use parotid sparing IMRT, they conceded that these "advanced" treatments are more costly to deliver compared to conventional techniques; essentially shifting costs from one pocket to the other. Whilst this may be true, the gains in the quality-adjusted life years as a result of parotid sparing IMRT cannot be ignored (304).

Although most of the literature addressing financial toxicity is dominated by research conducted in the United States, more work is needed in other regions to better understand the wider resource burdens placed on our workforce and society. This point is becoming increasingly relevant considering the improved survival outcomes and extended survivorship periods experienced by our patients. The integration of radiomics into routine clinical environments for the prediction of treatment related toxicities is one strategy to ameliorate this economic and emotional burden for our patients and workforce; however, this remains to be realised. Whilst technically challenging, the costs associated with radiomic analysis is likely to be inconsequential, considering software is freely available and analysis is performed on images already taken as part of standard clinical care. The work presented in this thesis has demonstrated the potential predictive power of radiomics, specifically for acute xerostomia and sticky saliva. Embracing radiomics in this regard will allow us to identify patients

who are at a higher risk of toxicity and may require additional supportive care during the most acute phase, allowing for a more patient centred and tailored approach to care. Current standard of care for treating the symptoms of xerostomia and sticky saliva includes pharmacological and non-pharmacological approaches (81, 305). Other novel approaches have been discussed in recent reviews on the topic such as gene therapy or stem cell transplant; however, their widespread use may be financially limiting (81, 89). In identifying high-risk patients, we could potentially direct these costly services more appropriately.

7.4 RECOMMENDATIONS FOR FUTURE WORK

Whilst this thesis was able to support the central hypothesis, future work is required to drive the field further forward. Work in this regard pertains to specific methodologies in the radiomics process as well as considerations that are relevant from a clinical perspective.

With respect to the radiomics process, there was variability in the methodology of some of the specific studies due to software changes and knowledge gained during this work. In particular, Study 1 used the default image intensity range of -1000 to 3000, and also used the default bin width of 10 HU. Whilst arguably, the generous intensity range would have had little impact on features extracted from the parotid gland tissues, the single large bin width explored limited a more complete comparison with results from the subsequent chapters which utilised the smaller bin width of 5 HU. Additionally, the inconsistency in LIFEx software versions used could be highlighted as a limitation; however, the decision to update the software as it became available was made to gain from the improvements (e.g., increased number of features extracted) with the latest versions.

There were several, more broad limitations in this thesis that extended beyond nuances of radiomics methodology. Radiation induced toxicity,

specifically xerostomia and the impact it has on the patient is subjective, and whilst not a limitation of this thesis per se, did present challenges. Treatment related toxicities can be captured using a range of methods using either patient reported outcome tools, patient questionnaires completed by the health care professional, and in the case of xerostomia and sticky saliva may include measuring salivary flow rate. Not only does this add a level of complexity and “noise”, but furthermore in this thesis, Study 3 and Study 4 used different scoring systems; although both are recognised and validated in clinical practice. Future work could consider integrating more objective tools to quantify radiation induced parotid gland damage, such as prostate specific membrane antigen - positron emission tomography (PSMA-PET). Whilst the salivary acini express PSMA, their capacity to exhibit this ligand uptake on PSMA is reduced following exposure to radiation (306). Considering radiomics software, specifically LIFEx, has been validated for use with PET imaging (123), assessment of parotid damage using such imaging could be an area of future research.

Whilst the parotid glands do produce 60%-65% of the total salivary output, the submandibular and sublingual glands also contribute to the total volume of saliva produced (307). This thesis examined the parotid glands exclusively; this decision was justified based on the limited capacity to spare the submandibular glands (specifically the contralateral gland) when treating the elective level II and III jugulodigastric nodes in the case of advanced stage oropharyngeal cancer (308). This is particularly relevant considering the majority of patients included in Study 3 had stage N2-3 disease. Furthermore, whilst others have included the submandibular glands in their radiomic driven predictive models, results have been mixed; and when assessing the generalisability of such models, Berger et al., (309) were unable to validate the predictive value of the mean dose to the submandibular glands established by van Dijk (207). In light of this, recent studies in this field continue to focus their attention exclusively on the parotid glands (310-312). In fact, Berger et al., (242) specifically examined the contralateral parotid gland to predict acute and late xerostomia in a cohort of 117 patients treated for a range of head and neck cancers. Their

decision to disregard the ipsilateral gland was justified by citing previous work demonstrating the superiority of the contralateral gland for xerostomia prediction (219, 239, 313). Despite these findings, the approach taken in Studies 3 and 4, i.e., to take the average value of each radiomic feature extracted from both parotid glands for each patient, can also be justified. Preliminary data analysis conducted in Study 3 indicated statistically similar radiomic feature values between each patient's ipsi and contralateral parotid glands. Furthermore, the mean dose received by each of the patients' glands in this study was also similar, giving grounds for examining the parotid glands in a combined fashion.

The promising results presented in this thesis considering the challenges acknowledged, allude to the need for future work to be done in this field. Validation of the predictive models generated in Study 3 is necessary. A strength of the model, however, is its statistical simplicity, whereby models generated on logistic regression have been found to perform as well or even better than more complex models based on neural networks which are at risk of overfitting due to the large number of input variables (246). Regardless, the decision not to split the cohort into a training and validation set was due to the insufficient patient numbers available; a requirement acknowledged in a systematic review of the application of radiomics quality scores in the literature (314). Even with 98 patients, data collected for Study 3 suffered from incomplete records available, a challenge also experienced by others (242). Combined with the listwise deletion approach adopted for the regression models the final patient numbers were limited and splitting the cohort at this point would have resulted in unacceptably small populations of the training and validation sets. A recent systematic review of the use of machine learning for the prediction of toxicity following treatment for head and neck cancer indicated that external validation of models trained is often not carried out, and collective efforts are needed from all radiomic researchers to overcome this shortcoming in the evidence base (210).

Irrespective of the complexity of the model, this final step is considered critical as a way of validating the model's predictive performance prior to clinical adoption, reflected in the heavy weight assigned to this step when calculating the radiomics quality score (107, 315). Generating a model that is both predictive based on the training set and generalizable when applied to unseen, independent data is not always successful. A prime example of this was when Burger et al., (309) assessed the generalizability of the models trained by van Dijk et al., (184). In this study they were unable to reproduce any of the significant associations between previously identified radiomic features with late xerostomia or sticky saliva. Whilst these results may appear disappointing, they simply emphasize once again the need for standardisation of radiomic research methodologies, along with the need for transparent reporting of such. Berger et al., (309) listed several differences between the studies; for example, scanners used for image acquisition. Limiting research to one or two specific scanners is not practical when working with big data common in multinational studies, and the use of ComBat harmonisation is one approach to overcome such challenges. Another specific difference between these two studies which may have influenced the results was the variation in voxel size. Based on the results of Study 2 of this thesis, this influence of spatial resampling on radiomic feature values is better understood, and some features are influenced less than others by this post-processing step. Standardising this variable can be handled simply through the use of a common isotropic voxel size. The work by Berger et al., (309) is a valuable contribution to the field of radiomics where negative results are infrequently reported (316).

Finally, Study 4 revealed the dynamic nature of radiomic features in response to radiation; an observation that was made with images acquired at just three timepoints. The prospect of leveraging data contained within *all* medical images taken for routine purposes motivates us to consider longitudinal radiomic analysis of cone-beam CTs (CBCT) taken during a course of treatment; an approach which has already been investigated in several disease sites, including head and neck cancer (317-319). Cone-beam CT is a common IGRT strategy used in head and neck radiotherapy,

with scans acquired weekly, if not daily (287). Whilst delta-radiomic analysis of these daily images would likely be too resource intensive, utilising these scans acquired on treatment rather than acquiring an additional simulation CT for radiomics would exploit the data they contain. With this approach to radiomic analysis, we could push forward the concept of “radiomic driven adaptive radiotherapy” in an offline environment. With this, the potential for radiomics to act as a clinical decision tool to support adaptive radiotherapy could be realised.

With the growing body of evidence utilising quantitative image analysis, along with the current narrative in this field, the next natural step to embed radiomics into clinical trials. A search of clinicaltrials.gov revealed 199 observational and 77 interventional studies containing the search term “radiomics” (320). Of these 276 studies, just 14 pertain to head and neck cancer. One RCT relevant to this thesis is investigating MRI - Guided Adaptive RadioTherapy for Reducing Xerostomia in Head and Neck Cancer (MARTHA-trial, Trial ID: NCT03972072). Although based on MRI, this trial is similar to Study 4 in that longitudinal assessment of radiomic features are quantified and correlated with late xerostomia. Furthermore, similarities exist with Study 3 in that xerostomia is also assessed using the EORTC QLQ-HN35 questionnaire. Whilst it is promising to see clinical trials such as this currently recruiting patients, the integration of radiomics into RTCs has been slow. This is largely attributed to the difficulties in standardising radiomic methodologies, inter-institutional variation of imaging protocols, and barriers to big data sharing (321); all challenges that were faced in the research undertaken for this thesis.

The use of radiomics to improve the prediction of acute xerostomia and sticky saliva can assist in the personalisation of radiotherapy by acting as a decision support tool. Prior to integrating radiomics into the clinical environment, the effect of methodological differences on radiomic features must be determined.

This thesis has identified radiomic features which are robust to specific methodologies in the radiomics pipeline; from variability in imaging acquisition and reconstruction parameters to delineation of the region of interest. These results have addressed previously unanswered questions relating to radiomic analysis of the parotid glands. Furthermore, this thesis has clearly demonstrated the improved predictive power of modelling parotid gland damage when radiomic features are combined with standard clinical and dosimetric endpoints. In doing so, this work has confirmed the superiority of including radiomics in order to overcome the historical challenges associated with estimating normal tissue complication probability.

The inclusion of radiomics in such a way will ultimately impact on the management of patients treated for head and neck cancer, which may include early intervention via adaptive radiotherapy or increased patient support. This novel approach to patient management is expected to reduce the personal and societal burden from the treatment of head and neck cancers.

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209-49.
2. Johnson DE, Burtneß B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers.* 2020;6(1):92.
3. Gormley M, Creaney G, Schache A, Ingarfield K, Conway DI. Reviewing the epidemiology of head and neck cancer: definitions, trends and risk factors. *Br Dent J.* 2022;233(9):780-6.
4. Ndon S, Singh A, Ha PK, Aswani J, Chan JY, Xu MJ. Human Papillomavirus-Associated Oropharyngeal Cancer: Global Epidemiology and Public Policy Implications. *Cancers (Basel).* 2023;15(16):4080.
5. Bossi P, Chan AT, Licitra L, Trama A, Orlandi E, Hui EP, et al. Nasopharyngeal carcinoma: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2021;32(4):452-65.
6. Conway DI, Purkayastha M, Chestnutt IG. The changing epidemiology of oral cancer: definitions, trends, and risk factors. *Br Dent J.* 2018;225(9):867-73.
7. Kreimer AR, Johansson M, Waterboer T, Kaaks R, Chang-Claude J, Drogen D, et al. Evaluation of human papillomavirus antibodies and risk of subsequent head and neck cancer. *J Clin Oncol.* 2013;31(21):2708-15.

8. Davis KS, Lim CM, Clump DA, Heron DE, Ohr JP, Kim S, et al. Tumor volume as a predictor of survival in HPV positive oropharyngeal cancer. *Head & neck*. 2016;38(Suppl 1):e1613-e7.
9. Spiotto MT, Taniguchi CM, Klopp AH, Colbert LE, Lin SH, Wang L, et al. Biology of the Radio- and Chemo-Responsiveness in HPV Malignancies. *Semin Radiat Oncol*. 2021;31(4):274-85.
10. Machiels JP, René Leemans C, Golusinski W, Grau C, Licitra L, Gregoire V, et al. Squamous cell carcinoma of the oral cavity, larynx, oropharynx and hypopharynx: EHNS-ESMO-ESTRO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2020;31(11):1462-75.
11. Parsons JT, Mendenhall WM, Stringer SP, Amdur RJ, Hinerman RW, Villaret DB, et al. Squamous cell carcinoma of the oropharynx: surgery, radiation therapy, or both. *Cancer*. 2002;94(11):2967-80.
12. Pedro C, Mira B, Silva P, Netto E, Pocinho R, Mota A, et al. Surgery vs. primary radiotherapy in early-stage oropharyngeal cancer. *Clin Transl Radiat Oncol*. 2018;9:18-22.
13. Morisod B, Simon C. Meta-analysis on survival of patients treated with transoral surgery versus radiotherapy for early-stage squamous cell carcinoma of the oropharynx. *Head Neck*. 2016;38 Suppl 1:e2143-e50.
14. Golusiński W. Functional Organ Preservation Surgery in Head and Neck Cancer: Transoral Robotic Surgery and Beyond. *Front Oncol*. 2019;9:293.
15. De Felice F, Cattaneo CG, Franco P. Radiotherapy and Systemic Therapies: Focus on Head and Neck Cancer. *Cancers (Basel)*. 2023;15(17).

16. Hay A, Nixon IJ. Recent advances in the understanding and management of oropharyngeal cancer. *F1000Research*. 2018;7:F1000 Faculty Rev-362.
17. Bernier J, Domette C, Ozsahin M, Matuszewska K, Lefèbvre J-L, Greiner RH, et al. Postoperative Irradiation with or without Concomitant Chemotherapy for Locally Advanced Head and Neck Cancer. *New England Journal of Medicine*. 2004;350(19):1945-52.
18. Cooper JS, Pajak TF, Forastiere AA, Jacobs J, Campbell BH, Saxman SB, et al. Postoperative Concurrent Radiotherapy and Chemotherapy for High-Risk Squamous-Cell Carcinoma of the Head and Neck. *New England Journal of Medicine*. 2004;350(19):1937-44.
19. Network NCC. Head and Neck Cancers. Version 3 2024 [Available from: https://www.nccn.org/professionals/physician_gls/pdf/head-and-neck.pdf].
20. Atwell D, Elks J, Cahill K, Hearn N, Vignarajah D, Lagopoulos J, et al. A Review of Modern Radiation Therapy Dose Escalation in Locally Advanced Head and Neck Cancer. *Clin Oncol (R Coll Radiol)*. 2020;32(5):330-41.
21. Lechner M, Liu J, Masterson L, Fenton TR. HPV-associated oropharyngeal cancer: epidemiology, molecular biology and clinical management. *Nat Rev Clin Oncol*. 2022;19(5):306-27.
22. Jiang L, Zhang Y, Yang Z, Liang F, Wu J, Wang R. A comparison of clinical outcomes between simultaneous integrated boost (SIB) versus sequential boost (SEQ) intensity modulated radiation therapy (IMRT) for head and neck cancer: A meta-analysis. *Medicine (Baltimore)*. 2019;98(34):e16942.

23. Jensen K, Friborg J, Hansen CR, Samsøe E, Johansen J, Andersen M, et al. The Danish Head and Neck Cancer Group (DAHANCA) 2020 radiotherapy guidelines. *Radiother Oncol.* 2020;151:149-51.
24. Overgaard J, Mohanti BK, Begum N, Ali R, Agarwal JP, Kuddu M, et al. Five versus six fractions of radiotherapy per week for squamous-cell carcinoma of the head and neck (IAEA-ACC study): A randomised, multicentre trial. *Lancet Oncol.* 2010;11(6):553-60.
25. Lacas B, Carmel A, Landais C, Wong SJ, Licitra L, Tobias JS, et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): An update on 107 randomized trials and 19,805 patients, on behalf of MACH-NC Group. *Radiother Oncol.* 2021;156:281-93.
26. Noronha V, Joshi A, Patil VM, Agarwal J, Ghosh-Laskar S, Budrukkar A, et al. Once-a-Week Versus Once-Every-3-Weeks Cisplatin Chemoradiation for Locally Advanced Head and Neck Cancer: A Phase III Randomized Noninferiority Trial. *J Clin Oncol.* 2018;36(11):1064-72.
27. Rades D, Zwaan I, Soror T, Idel C, Pries R, Bruchhage KL, et al. Chemoradiation with Cisplatin vs. Carboplatin for Squamous Cell Carcinoma of the Head and Neck (SCCHN). *Cancers (Basel).* 2023;15(13).
28. Bourhis J, Stein A, Paul de Boer J, Van Den Eynde M, Gold KA, Stintzing S, et al. Avelumab and cetuximab as a therapeutic combination: An overview of scientific rationale and current clinical trials in cancer. *Cancer Treat Rev.* 2021;97:102172.
29. Gillison ML, Trotti AM, Harris J, Eisbruch A, Harari PM, Adelstein DJ, et al. Radiotherapy plus cetuximab or cisplatin in human papillomavirus-positive oropharyngeal cancer (NRG Oncology RTOG 1016): a randomised, multicentre, non-inferiority trial. *Lancet.* 2019;393(10166):40-50.

30. Tao Y, Aupérin A, Sun X, Sire C, Martin L, Coutte A, et al. Avelumab-cetuximab-radiotherapy versus standards of care in locally advanced squamous-cell carcinoma of the head and neck: The safety phase of a randomised phase III trial GORTEC 2017-01 (REACH). *Eur J Cancer*. 2020;141:21-9.
31. Collins JM, Gulley JL. Product review: avelumab, an anti-PD-L1 antibody. *Hum Vaccin Immunother*. 2019;15(4):891-908.
32. Cou GORTe. Randomized Trial of Avelumab-cetuximab-radiotherapy Versus SOC in LA SCCHN (REACH) (REACH) 2023 [updated 2023-06-22. Available from: <https://clinicaltrials.gov/study/NCT02999087>.
33. Hill RM, Rocha S, Parsons JL. Overcoming the Impact of Hypoxia in Driving Radiotherapy Resistance in Head and Neck Squamous Cell Carcinoma. *Cancers (Basel)*. 2022;14(17).
34. Overgaard J. Hypoxic modification of radiotherapy in squamous cell carcinoma of the head and neck – A systematic review and meta-analysis. *Radiotherapy and Oncology*. 2011;100(1):22-32.
35. Lassen P, Eriksen JG, Hamilton-Dutoit S, Tramm T, Alsner J, Overgaard J, et al. HPV-associated p16-expression and response to hypoxic modification of radiotherapy in head and neck cancer. *Radiother Oncol*. 2010;94(1):30-5.
36. Sung SY, Kim YS, Kim SH, Lee SJ, Lee SW, Kwak YK. Current Evidence of a Deintensification Strategy for Patients with HPV-Related Oropharyngeal Cancer. *Cancers (Basel)*. 2022;14(16).
37. Yang X, Tridandapani S, Beitler JJ, Yu DS, Chen Z, Kim S, et al. Diagnostic accuracy of ultrasonic histogram features to evaluate radiation

toxicity of the parotid glands: a clinical study of xerostomia following head-and-neck cancer radiotherapy. *Acad Radiol*. 2014;21(10):1304-13.

38. Wang K, Tepper JE. Radiation therapy-associated toxicity: Etiology, management, and prevention. *CA Cancer J Clin*. 2021;71(5):437-54.

39. Rades D, Warwas B, Cremers F, Gerull K, Pries R, Leichtle A, et al. The First Prognostic Tool to Estimate the Risk of Late Grade ≥ 3 Xerostomia in Patients Irradiated for Head-and-Neck Cancer. *Anticancer Res*. 2022;42(6):3035-9.

40. Men K, Geng H, Zhong H, Fan Y, Lin A, Xiao Y. A Deep Learning Model for Predicting Xerostomia Due to Radiation Therapy for Head and Neck Squamous Cell Carcinoma in the RTOG 0522 Clinical Trial. *Int J Radiat Oncol Biol Phys*. 2019;105(2):440-7.

41. De Felice F, Scarabelli MG, De Pietro R, Chiarello G, Di Giammarco F, Cattaneo CG, et al. Relationship between Salivary Amylase and Xerostomia in Intensity-Modulated Radiation Therapy for Head and Neck Cancer: A Prospective Pilot Study. *Curr Oncol*. 2022;29(9):6564-72.

42. Kawashima M, Kawabata T, Ando C, Sakuma M, Aoyama T, Ogawa H, et al. Radiation-induced xerostomia and cariogenic dietary habits. *Support Care Cancer*. 2024;32(2):92.

43. Jasmer KJ, Gilman KE, Muñoz Forti K, Weisman GA, Limesand KH. Radiation-Induced Salivary Gland Dysfunction: Mechanisms, Therapeutics and Future Directions. *J Clin Med*. 2020;9(12).

44. Porcheri C, Mitsiadis TA. Physiology, Pathology and Regeneration of Salivary Glands. *Cells*. 2019;8(9).

45. Chason HM, Downs BW. Anatomy, Head and Neck, Parotid Gland. In: StatPearls. Treasure Island (FL): Stats Pearl Publishing; October 24, 2022.
46. Iorgulescu G. Saliva between normal and pathological. Important factors in determining systemic and oral health. *J Med Life*. 2009;2(3):303-7.
47. Deasy JO, Moiseenko V, Marks L, Chao KS, Nam J, Eisbruch A. Radiotherapy dose-volume effects on salivary gland function. *Int J Radiat Oncol Biol Phys*. 2010;76(3 Suppl):S58-63.
48. Eisbruch A, Rhodus N, Rosenthal D, Murphy B, Rasch C, Sonis S, et al. How should we measure and report radiotherapy-induced xerostomia? *Semin Radiat Oncol*. 2003;13(3):226-34.
49. Vissink A, Jansma J, Spijkervet FK, Burlage FR, Coppes RP. Oral sequelae of head and neck radiotherapy. *Crit Rev Oral Biol Med*. 2003;14(3):199-212.
50. Winter C, Keimel R, Gugatschka M, Kolb D, Leitinger G, Roblegg E. Investigation of Changes in Saliva in Radiotherapy-Induced Head Neck Cancer Patients. *Int J Environ Res Public Health*. 2021;18(4).
51. Burlage FR, Coppes RP, Meertens H, Stokman MA, Vissink A. Parotid and submandibular/sublingual salivary flow during high dose radiotherapy. *Radiother Oncol*. 2001;61(3):271-4.
52. Murdoch-Kinch CA, Kim HM, Vineberg KA, Ship JA, Eisbruch A. Dose-effect relationships for the submandibular salivary glands and implications for their sparing by intensity modulated radiotherapy. *Int J Radiat Oncol Biol Phys*. 2008;72(2):373-82.

53. Lombaert I, Movahednia MM, Adine C, Ferreira JN. Concise Review: Salivary Gland Regeneration: Therapeutic Approaches from Stem Cells to Tissue Organoids. *Stem Cells*. 2017;35(1):97-105.
54. Han P, Lakshminarayanan P, Jiang W, Shpitser I, Hui X, Lee SH, et al. Dose/Volume histogram patterns in Salivary Gland subvolumes influence xerostomia injury and recovery. *Sci Rep*. 2019;9(1):3616.
55. van Luijk P, Pringle S, Deasy JO, Moiseenko VV, Faber H, Hovan A, et al. Sparing the region of the salivary gland containing stem cells preserves saliva production after radiotherapy for head and neck cancer. *Sci Transl Med*. 2015;7(305):305ra147.
56. Steenbakkers R, Stokman M, Kierkels RGJ, Schuurman M, Hoek AVD, Bijl H, et al., editors. Stem Cell Sparing IMRT for Head and Neck Cancer Patients: A Double-Blind Randomized Controlled Trial. *ASTRO*; 2019.
57. Common Terminology Criteria for Adverse Events (CTCAE) Version 5. US Department of Health and Human Services, National Institutes of Health, National Cancer Institute; 2017. Available from: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm#ctc_50
58. Cheng SC, Wu VW, Kwong DL, Ying MT. Assessment of post-radiotherapy salivary glands. *Br J Radiol*. 2011;84(1001):393-402.
59. Li Y, Li X, Pang R, Yang G, Tian M, Zhao T, et al. Diagnosis, Prevention, and Treatment of Radiotherapy-Induced Xerostomia: A Review. *J Oncol*. 2022;2022:7802334.

60. Pow EH, Kwong DL, McMillan AS, Wong MC, Sham JS, Leung LH, et al. Xerostomia and quality of life after intensity-modulated radiotherapy vs. conventional radiotherapy for early-stage nasopharyngeal carcinoma: initial report on a randomized controlled clinical trial. *Int J Radiat Oncol Biol Phys.* 2006;66(4):981-91.
61. Eisbruch A, Kim HM, Terrell JE, Marsh LH, Dawson LA, Ship JA. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. *Int J Radiat Oncol Biol Phys.* 2001;50(3):695-704.
62. Wu VWC, Leung KY. A Review on the Assessment of Radiation Induced Salivary Gland Damage After Radiotherapy. *Front Oncol.* 2019;9:1090.
63. Ishikawa S, Sugimoto M, Kitabatake K, Tu M, Sugano A, Yamamori I, et al. Effect of timing of collection of salivary metabolomic biomarkers on oral cancer detection. *Amino Acids.* 2017;49(4):761-70.
64. Murdoch-Kinch CA, Russo N, Griffith S, Braun T, Eisbruch A, D'Silva NJ. Recovery of salivary epidermal growth factor in parotid saliva following parotid sparing radiation therapy: a proof-of-principle study. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 2011;111(1):64-70.
65. Ying M, Wu VW, Kwong DL. Comparison of sonographic appearance of normal and postradiotherapy parotid glands: a preliminary study. *Ultrasound Med Biol.* 2007;33(8):1244-50.
66. Nömayr A, Lell M, Sweeney R, Bautz W, Lukas P. MRI appearance of radiation-induced changes of normal cervical tissues. *Eur Radiol.* 2001;11(9):1807-17.
67. Simo R, Homer J, Clarke P, Mackenzie K, Paleri V, Pracy P, et al. Follow-up after treatment for head and neck cancer: United Kingdom

National Multidisciplinary Guidelines. *J Laryngol Otol.* 2016;130(S2):S208-S11.

68. Duffton A, Kemp O, Devlin L, Hay L, McLoone P, Paterson C. Feasibility of DW-MRI analysis of salivary glands during head and neck radiotherapy. *Tech Innov Patient Support Radiat Oncol.* 2021;19:46-51.

69. Mohan V, Bruin NM, van de Kamer JB, Sonke JJ, Vogel WV. The increasing potential of nuclear medicine imaging for the evaluation and reduction of normal tissue toxicity from radiation treatments. *Eur J Nucl Med Mol Imaging.* 2021;48(12):3762-75.

70. Buus S, Grau C, Munk OL, Bender D, Jensen K, Keiding S. 11C-methionine PET, a novel method for measuring regional salivary gland function after radiotherapy of head and neck cancer. *Radiother Oncol.* 2004;73(3):289-96.

71. Palmer MK. WHO Handbook for Reporting Results of Cancer Treatment. *Br J Cancer* 1982;45(3):484–5.

72. Trotti A, Byhardt R, Stetz J, Gwede C, Corn B, Fu K, et al. Common toxicity criteria: version 2.0. an improved reference for grading the acute effects of cancer treatment: impact on radiotherapy. *Int J Radiat Oncol Biol Phys.* 2000;47(1):13-47.

73. Cox JD, Stetz J, Pajak TF. Toxicity criteria of the Radiation Therapy Oncology Group (RTOG) and the European Organization for Research and Treatment of Cancer (EORTC). *Int J Radiat Oncol Biol Phys.* 1995;31(5):1341-6.

74. Jensen K, Bonde Jensen A, Grau C. The relationship between observer-based toxicity scoring and patient assessed symptom severity after treatment for head and neck cancer. A correlative cross sectional

study of the DAHANCA toxicity scoring system and the EORTC quality of life questionnaires. *Radiother Oncol.* 2006;78(3):298-305.

75. Laugsand EA, Sprangers MA, Bjordal K, Skorpen F, Kaasa S, Klepstad P. Health care providers underestimate symptom intensities of cancer patients: a multicenter European study. *Health Qual Life Outcomes.* 2010;8:104.

76. National Cancer Institute PRO-CTCAE [Available from: <https://healthcaredelivery.cancer.gov/pro-ctcae/>].

77. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *J Natl Cancer Inst.* 2014;106(9).

78. Heutte N, Plisson L, Lange M, Prevost V, Babin E. Quality of life tools in head and neck oncology. *Eur Ann Otorhinolaryngol Head Neck Dis.* 2014;131(1):33-47.

79. Silveira AP, Gonçalves J, Sequeira T, Ribeiro C, Lopes C, Monteiro E, et al. Patient reported outcomes in head and neck cancer: selecting instruments for quality of life integration in clinical protocols. *Head Neck Oncol.* 2010;2:32.

80. Holländer-Mieritz C, Johansen J, Taarnhøj GA, Johansen C, Vogelius IR, Kristensen CA, et al. Systematic use of patient reported outcome during radiotherapy for head and neck cancer: study protocol for the national DAHANCA 38 trial. *Acta Oncol.* 2020;59(5):603-7.

81. Nathan CO, Asarkar AA, Entezami P, Corry J, Stojan P, Poorten VV, et al. Current management of xerostomia in head and neck cancer patients. *Am J Otolaryngol.* 2023;44(4):103867.

82. Koukourakis MI. Amifostine: is there evidence of tumor protection? *Semin Oncol.* 2003;30(6 Suppl 18):18-30.
83. Gu J, Zhu S, Li X, Wu H, Li Y, Hua F. Effect of amifostine in head and neck cancer patients treated with radiotherapy: a systematic review and meta-analysis based on randomized controlled trials. *PLoS One.* 2014;9(5):e95968.
84. Riley P, Glenny AM, Hua F, Worthington HV. Pharmacological interventions for preventing dry mouth and salivary gland dysfunction following radiotherapy. *Cochrane Database Syst Rev.* 2017;7(7):CD012744.
85. Sood AJ, Fox NF, O'Connell BP, Lovelace TL, Nguyen SA, Sharma AK, et al. Salivary gland transfer to prevent radiation-induced xerostomia: a systematic review and meta-analysis. *Oral Oncol.* 2014;50(2):77-83.
86. DE Felice F, Pranno N, Papi P, Brugnoletti O, Tombolini V, Polimeni A. Xerostomia and Clinical Outcomes in Definitive Intensity Modulated Radiotherapy (IMRT). *In Vivo.* 2020;34(2):623-9.
87. Gupta T, Kannan S, Ghosh-Laskar S, Agarwal JP. Systematic review and meta-analyses of intensity-modulated radiation therapy versus conventional two-dimensional and/or or three-dimensional radiotherapy in curative-intent management of head and neck squamous cell carcinoma. *PLoS One.* 2018;13(7):e0200137.
88. Gasinska A, Fowler JF, Lind BK, Urbanski K. Influence of overall treatment time and radiobiological parameters on biologically effective doses in cervical cancer patients treated with radiation therapy alone. *Acta Oncol.* 2004;43(7):657-66.

89. Sasportas LS, Hosford DN, Sodini MA, Waters DJ, Zambricki EA, Barral JK, et al. Cost-effectiveness landscape analysis of treatments addressing xerostomia in patients receiving head and neck radiation therapy. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2013;116(1):e37-51.
90. Fox NF, Xiao C, Sood AJ, Lovelace TL, Nguyen SA, Sharma A, et al. Hyperbaric oxygen therapy for the treatment of radiation-induced xerostomia: a systematic review. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2015;120(1):22-8.
91. Warde P, Kroll B, O'Sullivan B, Aslanidis J, Tew-George E, Waldron J, et al. A phase II study of Biotene in the treatment of postradiation xerostomia in patients with head and neck cancer. *Support Care Cancer*. 2000;8(3):203-8.
92. Porangaba LP, de Melo Garcia F, Rabelo APAA, Andrade AP, de Abreu Alves F, Pellizzon ACA, et al. Randomized Double-Blind Placebo-Controlled Study of Salivary Substitute with Enzymatic System for Xerostomia in Patients Irradiated in Head and Neck Region. *Curr Oncol*. 2024;31(2):1102-12.
93. Furness S, Worthington HV, Bryan G, Birchenough S, McMillan R. Interventions for the management of dry mouth: topical therapies. *Cochrane Database Syst Rev*. 2011(12):CD008934.
94. Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constone LS, Eisbruch A, et al. Use of normal tissue complication probability models in the clinic. *Int J Radiat Oncol Biol Phys*. 2010;76(3 Suppl):S10-9.
95. Tambas M, van der Laan HP, Steenbakkers RJHM, Doyen J, Timmermann B, Orlandi E, et al. Current practice in proton therapy delivery in adult cancer patients across Europe. *Radiother Oncol*. 2022;167:7-13.

96. Jackson A, Marks LB, Bentzen SM, Eisbruch A, Yorke ED, Ten Haken RK, et al. The lessons of QUANTEC: recommendations for reporting and gathering data on dose-volume dependencies of treatment outcome. *Int J Radiat Oncol Biol Phys.* 2010;76(3 Suppl):S155-60.
97. Brodin NP, Tomé WA. Revisiting the dose constraints for head and neck OARs in the current era of IMRT. *Oral Oncol.* 2018;86:8-18.
98. Hansen CR, Bertelsen A, Zukauskaite R, Johnsen L, Bernchou U, Thwaites DJ, et al. Prediction of radiation-induced mucositis of H&N cancer patients based on a large patient cohort. *Radiother Oncol.* 2020;147:15-21.
99. Lee TF, Chao PJ, Wang HY, Hsu HC, Chang P, Chen WC. Normal tissue complication probability model parameter estimation for xerostomia in head and neck cancer patients based on scintigraphy and quality of life assessments. *BMC Cancer.* 2012;12:567.
100. Allen Li X, Alber M, Deasy JO, Jackson A, Ken Jee KW, Marks LB, et al. The use and QA of biologically related models for treatment planning: short report of the TG-166 of the therapy physics committee of the AAPM. *Med Phys.* 2012;39(3):1386-409.
101. Dijkema T, Raaijmakers CP, Ten Haken RK, Roesink JM, Braam PM, Houweling AC, et al. Parotid gland function after radiotherapy: the combined michigan and utrecht experience. *Int J Radiat Oncol Biol Phys.* 2010;78(2):449-53.
102. Isaksson LJ, Pepa M, Zaffaroni M, Marvaso G, Alterio D, Volpe S, et al. Machine Learning-Based Models for Prediction of Toxicity Outcomes in Radiotherapy. *Front Oncol.* 2020;10:790.

103. Stieb S, Lee A, van Dijk LV, Frank S, Fuller CD, Blanchard P. NTCP Modeling of Late Effects for Head and Neck Cancer: A Systematic Review. *Int J Part Ther*. 2021;8(1):95-107.
104. Tommasino F, Nahum A, Cella L. Increasing the power of tissue control and normal tissue complication probability modelling in radiotherapy: recent trends and current issues. *Translational Cancer Research*. 2017;6:807-21.
105. Measurements ICoRUa. ICRU Report 62. Prescribing, recording, and reporting photon beam therapy (Supplement to ICRU Report 50). Bethesda, MD1999.
106. Brodin NP, Kabarriti R, Garg MK, Guha C, Tomé WA. Systematic Review of Normal Tissue Complication Models Relevant to Standard Fractionation Radiation Therapy of the Head and Neck Region Published After the QUANTEC Reports. *Int J Radiat Oncol Biol Phys*. 2018;100(2):391-407.
107. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-62.
108. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to Radiomics. *J Nucl Med*. 2020;61(4):488-95.
109. Kang W, Qiu X, Luo Y, Luo J, Liu Y, Xi J, et al. Application of radiomics-based multiomics combinations in the tumor microenvironment and cancer prognosis. *J Transl Med*. 2023;21(1):598.
110. Tortora M, Gemini L, Scaravilli A, Ugga L, Ponsiglione A, Stanzione A, et al. Radiomics Applications in Head and Neck Tumor Imaging: A Narrative Review. *Cancers (Basel)*. 2023;15(4).

111. Bagher-Ebadian H, Lu M, Siddiqui F, Ghanem AI, Wen N, Wu Q, et al. Application of radiomics for the prediction of HPV status for patients with head and neck cancers. *Med Phys*. 2020;47(2):563-75.
112. Spadarella G, Ugga L, Calareso G, Villa R, D'Aniello S, Cuocolo R. The impact of radiomics for human papillomavirus status prediction in oropharyngeal cancer: systematic review and radiomics quality score assessment. *Neuroradiology*. 2022;64(8):1639-47.
113. Leijenaar RT, Bogowicz M, Jochems A, Hoebers FJ, Wesseling FW, Huang SH, et al. Development and validation of a radiomic signature to predict HPV (p16) status from standard CT imaging: a multicenter study. *Br J Radiol*. 2018;91(1086):20170498.
114. Bogowicz M, Riesterer O, Ikenberg K, Stieb S, Moch H, Studer G, et al. Computed Tomography Radiomics Predicts HPV Status and Local Tumor Control After Definitive Radiochemotherapy in Head and Neck Squamous Cell Carcinoma. *Int J Radiat Oncol Biol Phys*. 2017;99(4):921-8.
115. Barua S, Elhalawani H, Volpe S, Al Feghali KA, Yang P, Ng SP, et al. Computed Tomography Radiomics Kinetics as Early Imaging Correlates of Osteoradionecrosis in Oropharyngeal Cancer Patients. *Front Artif Intell*. 2021;4:618469.
116. Sheikh K, Lee SH, Cheng Z, Lakshminarayanan P, Peng L, Han P, et al. Predicting acute radiation induced xerostomia in head and neck Cancer using MR and CT Radiomics of parotid and submandibular glands. *Radiat Oncol*. 2019;14(1):131.
117. Shur JD, Doran SJ, Kumar S, Ap Dafydd D, Downey K, O'Connor JPB, et al. Radiomics in Oncology: A Practical Guide. *Radiographics*. 2021;41(6):1717-32.

118. Lopez CJ, Nagornaya N, Parra NA, Kwon D, Ishkanian F, Markoe AM, et al. Association of Radiomics and Metabolic Tumor Volumes in Radiation Treatment of Glioblastoma Multiforme. *International Journal of Radiation Oncology*Biography*Physics*. 2017;97(3):586-95.
119. Lian C, Ruan S, Denœux T, Jardin F, Vera P. Selecting radiomic features from FDG-PET images for cancer treatment outcome prediction. *Medical Image Analysis*. 2016;32:257-68.
120. Huynh E, Coroller TP, Narayan V, Agrawal V, Hou Y, Romano J, et al. CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer. *Radiotherapy and Oncology*. 2016;120(2):258-66.
121. Fave X, Mackin D, Yang J, Zhang J, Fried D, Balter P, et al. Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer? *Medical physics*. 2015;42(12):6784-97.
122. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res*. 2017;77(21):e104-e7.
123. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Res*. 2018;78(16):4786-9.
124. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys*. 2015;42(3):1341-53.

125. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*. 2020;295(2):328-38.
126. Bibault JE, Xing L, Giraud P, El Ayachy R, Giraud N, Decazes P, et al. Radiomics: A primer for the radiation oncologist. *Cancer Radiother*. 2020;24(5):403-10.
127. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563-77.
128. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp*. 2018;2(1):36.
129. Nardone V, Tini P, Nioche C, Mazzei MA, Carfagno T, Battaglia G, et al. Texture analysis as a predictor of radiation-induced xerostomia in head and neck patients undergoing IMRT. *Radiol Med*. 2018;123(6):415-23.
130. The Image Biomarker Standardisation Initiative [cited 2024 March 4th]. Available from: https://ibsi.readthedocs.io/en/latest/03_Image_features.html.
131. Gillies R KP, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology*. 2016;278(2):563-77.
132. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man, and Cybernetics*. 1989;15(5):1264-74.

133. Nioche C, Orlhac F, Buvat I. Local Image Features Extraction - Texture Users Guide. 2019. Available from: <https://www.lifexsoft.org/images/phocagallery/documentation/LIFEx/UserGuide/LIFExUserGuide.pdf>
134. Wong AJ, Kanwar A, Mohamed AS, Fuller CD. Radiomics in head and neck cancer: from exploration to application. *Transl Cancer Res.* 2016;5(4):371-82.
135. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging-"how-to" guide and critical reflection. *Insights Imaging.* 2020;11(1):91.
136. Reiazi R, Abbas E, Famiyeh P, Rezaie A, Kwan JYY, Patel T, et al. The impact of the variation of imaging parameters on the robustness of Computed Tomography radiomic features: A review. *Comput Biol Med.* 2021;133:104400.
137. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int J Radiat Oncol Biol Phys.* 2018;102(4):1143-58.
138. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging.* 2013;26(6):1045-57.
139. Zuley M, Jarosz R, Kirk S, Colen R, K G, N A. Radiology Data from The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma [TCGA-HNSC] collection The Cancer Imaging Archive 2020
140. Head-and-neck squamous cell carcinoma patients with CT taken during pre-treatment, mid-treatment, and post-treatment Dataset

. [Internet]. 2018. Available from:

<https://www.cancerimagingarchive.net/collection/hnsc-3dct-rt/>

141. Eriksen JG, Salembier C, Rivera S, De Bari B, Berger D, Mantello G, et al. Four years with FALCON - an ESTRO educational project: achievements and perspectives. *Radiother Oncol*. 2014;112(1):145-9.

142. Brouwer CL, Steenbakkers RJ, Bourhis J, Budach W, Grau C, Grégoire V, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother Oncol*. 2015;117(1):83-90.

143. Sherman AC, Simonton S, Adams DC, Vural E, Owens B, Hanna E. Assessing quality of life in patients with head and neck cancer: cross-validation of the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Head and Neck module (QLQ-H&N35). *Arch Otolaryngol Head Neck Surg*. 2000;126(4):459-67.

144. Scalco E, Fiorino C, Cattaneo GM, Sanguineti G, Rizzo G. Texture analysis for the assessment of structural changes in parotid glands induced by radiotherapy. *Radiother Oncol*. 2013;109(3):384-7.

145. Riegel AC, Berson AM, Destian S, Ng T, Tena LB, Mitnick RJ, et al. Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion. *Int J Radiat Oncol Biol Phys*. 2006;65(3):726-32.

146. Breen SL, Publicover J, De Silva S, Pond G, Brock K, O'Sullivan B, et al. Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers. *Int J Radiat Oncol Biol Phys*. 2007;68(3):763-70.

147. Mukesh M, Benson R, Jena R, Hoole A, Roques T, Scrase C, et al. Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help? *Br J Radiol.* 2012;85(1016):e530-6.
148. Bekelman JE, Wolden S, Lee N. Head-and-neck target delineation among radiation oncology residents after a teaching intervention: a prospective, blinded pilot study. *Int J Radiat Oncol Biol Phys.* 2009;73(2):416-23.
149. Loo SW, Martin WM, Smith P, Cherian S, Roques TW. Interobserver variation in parotid gland delineation: a study of its impact on intensity-modulated radiotherapy solutions with a systematic review of the literature. *Br J Radiol.* 2012;85(1016):1070-7.
150. Geets X, Daisne JF, Arcangeli S, Coche E, De Poel M, Duprez T, et al. Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI. *Radiother Oncol.* 2005;77(1):25-31.
151. Nelms BE, Tomé WA, Robinson G, Wheeler J. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. *Int J Radiat Oncol Biol Phys.* 2012;82(1):368-78.
152. Liu C, Kong X, Gong G, Liu T, Li B, Yin Y. Error in the parotid contour delineated using computed tomography images rather than magnetic resonance images during radiotherapy planning for nasopharyngeal carcinoma. *Jpn J Radiol.* 2014;32(4):211-6.
153. Dewas S, Bibault JE, Blanchard P, Vautravers-Dewas C, Pointreau Y, Denis F, et al. Delineation in thoracic oncology: a prospective study of the effect of training on contour variability and dosimetric consequences. *Radiat Oncol.* 2011;6:118.

154. Vinod SK, Min M, Jameson MG, Holloway LC. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. *J Med Imaging Radiat Oncol*. 2016;60(3):393-406.
155. Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. *J Clin Oncol*. 2010;28(18):2996-3001.
156. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in Auto-Segmentation. *Semin Radiat Oncol*. 2019;29(3):185-97.
157. Owens CA, Peterson CB, Tang C, Koay EJ, Yu W, Mackin DS, et al. Lung tumor segmentation methods: Impact on the uncertainty of radiomics features for non-small cell lung cancer. *PLoS One*. 2018;13(10):e0205003.
158. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One*. 2014;9(7):e102107.
159. Mori M, Benedetti G, Partelli S, Sini C, Andreasi V, Broggi S, et al. Ct radiomic features of pancreatic neuroendocrine neoplasms (panNEN) are robust against delineation uncertainty. *Phys Med*. 2019;57:41-6.
160. Belli ML, Mori M, Broggi S, Cattaneo GM, Bettinardi V, Dell'Oca I, et al. Quantifying the robustness of [Phys Med. 2018;49:105-11.
161. Pavic M, Bogowicz M, Würms X, Glatz S, Finazzi T, Riesterer O, et al. Influence of inter-observer delineation variability on radiomics stability in different tumor sites. *Acta Oncol*. 2018;57(8):1070-4.

162. Huang Q, Lu L, Dercle L, Lichtenstein P, Li Y, Yin Q, et al. Interobserver variability in tumor contouring affects the use of radiomics to predict mutational status. *J Med Imaging (Bellingham)*. 2018;5(1):011005.
163. Haga A, Takahashi W, Aoki S, Nawa K, Yamashita H, Abe O, et al. Classification of early stage non-small cell lung cancers on computed tomographic images into histological types using radiomic features: interobserver delineation variability analysis. *Radiol Phys Technol*. 2018;11(1):27-35.
164. Zou KH, Warfield SK, Bharatha A, Tempany CM, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol*. 2004;11(2):178-89.
165. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*. 2012;30(9):1323-41.
166. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol*. 2018;19(9):1180-91.
167. Chou K-T, Latifi K, Moros E, Feygelman V, Huang T-C, Dilling T, et al. Evaluation of Radiomics Features Stability When Deformable Image Registration is Applied. *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies*. 2018;2:153-8.
168. Awan M, Kalpathy-Cramer J, Gunn GB, Beadle BM, Garden AS, Phan J, et al. Prospective assessment of an atlas-based intervention

combined with real-time software feedback in contouring lymph node levels and organs-at-risk in the head and neck: Quantitative assessment of conformance to expert delineation. *Pract Radiat Oncol*. 2013;3(3):186-93.

169. Walker GV, Awan M, Tao R, Koay EJ, Boehling NS, Grant JD, et al. Prospective randomized double-blind study of atlas-based organ-at-risk autosegmentation-assisted radiation planning in head and neck cancer. *Radiother Oncol*. 2014;112(3):321-5.

170. Hänsch A, Schwier M, Gass T, Morgas T, Haas B, Dicken V, et al. Evaluation of deep learning methods for parotid gland segmentation from CT images. *J Med Imaging (Bellingham)*. 2019;6(1):011005.

171. Leijenaar RT, Carvalho S, Velazquez ER, van Elmpst WJ, Parmar C, Hoekstra OS, et al. Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta Oncol*. 2013;52(7):1391-7.

172. ProKnow. How are contouring accuracy “scores” calculated? 2016 [Available from: <https://blog.proknowsystems.com/anatomy-contouring/how-are-contouring-accuracy-scores-calculated/>].

173. Delaney AR, Dahele M, Slotman BJ, Verbakel WFAR. Is accurate contouring of salivary and swallowing structures necessary to spare them in head and neck VMAT plans? *Radiother Oncol*. 2018;127(2):190-6.

174. Bagher-Ebadian H, Siddiqui F, Liu C, Movsas B, Chetty IJ. On the impact of smoothing and noise on robustness of CT and CBCT radiomics features for patients with head and neck cancers. *Med Phys*. 2017;44(5):1755-70.

175. Galavis P, Hollensen C, Jallow N, et al. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncologica*. 2010;49:1012-6.
176. Yip S, McCall K, Aristophanous M, Chen AB, Aerts HJ, Berbeco R. Comparison of texture features derived from static and respiratory-gated PET images in non-small cell lung cancer. *PLoS One*. 2014;9(12):e115510.
177. Pota M, Scalco E, Sanguineti G, Farneti A, Cattaneo GM, Rizzo G, et al. Early prediction of radiotherapy-induced parotid shrinkage and toxicity based on CT radiomics and fuzzy classification. *Artif Intell Med*. 2017;81:41-53.
178. Liu Y, Shi H, Huang S, Chen X, Zhou H, Chang H, et al. Early prediction of acute xerostomia during radiation therapy for nasopharyngeal cancer based on delta radiomics from CT images. *Quant Imaging Med Surg*. 2019;9(7):1288-302.
179. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys*. 2017;44(3):1050-62.
180. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring Computed Tomography Scanner Variability of Radiomics Features. *Invest Radiol*. 2015;50(11):757-65.
181. Larue R. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta Oncologica*. 2017;56(11):1544-53.
182. Haga A, Takahashi W, Aoki S, Nawa K, Yamashita H, Abe O, et al. Standardization of imaging features for radiomics analysis. *J Med Invest*. 2019;66(1.2):35-7.

183. Lu L, Ehmke RC, Schwartz LH, Zhao B. Assessing Agreement between Radiomic Features Computed for Multiple CT Imaging Settings. *PLoS One*. 2016;11(12):e0166550.
184. van Dijk LV, Brouwer CL, van der Schaaf A, Burgerhof JGM, Beukinga RJ, Langendijk JA, et al. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. *Radiother Oncol*. 2017;122(2):185-91.
185. Gabryś HS, Buettner F, Sterzing F, Hauswald H, Bangert M. Design and Selection of Machine Learning Methods Using Radiomics and Dosiomics for Normal Tissue Complication Probability Modeling of Xerostomia. *Front Oncol*. 2018;8:35.
186. Zwanenburg A, Leger S, Vallières M. Image Biomarker Standardisation Initiative. 2016.
187. Mackin D, Fave X, Zhang L, Yang J, Jones AK, Ng CS, et al. Harmonizing the pixel size in retrospective computed tomography radiomics studies. *PLoS One*. 2017;12(9):e0178524.
188. Ibrahim A, Refaee T, Primakov S, Barufaldi B, Acciavatti RJ, Granzier RWY, et al. The Effects of In-Plane Spatial Resolution on CT-Based Radiomic Features' Stability with and without ComBat Harmonization. *Cancers (Basel)*. 2021;13(8).
189. Mottola M, Ursprung S, Rundo L, Sanchez LE, Klatte T, Mendichovszky I, et al. Reproducibility of CT-based radiomic features against image resampling and perturbations for tumour and healthy kidney in renal cancer patients. *Sci Rep*. 2021;11(1):11542.

190. Bogowicz M, Riesterer O, Bundschuh RA, Veit-Haibach P, Hüllner M, Studer G, et al. Stability of radiomic features in CT perfusion maps. *Phys Med Biol*. 2016;61(24):8736-49.
191. Ger RB, Zhou S, Chi PM, Lee HJ, Layman RR, Jones AK, et al. Comprehensive Investigation on Controlling for CT Imaging Variabilities in Radiomics Studies. *Sci Rep*. 2018;8(1):13047.
192. Thor M, Tyagi N, Hatzoglou V, Apte A, Saleh Z, Riaz N, et al. A Magnetic Resonance Imaging-based approach to quantify radiation-induced normal tissue injuries applied to trismus in head and neck cancer. *Phys Imaging Radiat Oncol*. 2017;1:34-40.
193. Nardone V, Tini P, Nioche C, Biondi M, Sebaste L, Mazzei MA, et al. Texture analysis of parotid gland as a predictive factor of radiation induced xerostomia: A subset analysis. *Radiother Oncol*. 2017;122(2):321.
194. van Dijk LV, Langendijk JA, Sijtsema NM, Steenbakkers RJHM. Reply letter to "Texture analysis of parotid gland as a predictive factor of radiation induced xerostomia: A subset analysis". *Radiother Oncol*. 2017;122(2):322.
195. Bruixola G, Remacha E, Jiménez-Pastor A, Dualde D, Viala A, Montón JV, et al. Radiomics and radiogenomics in head and neck squamous cell carcinoma: Potential contribution to patient management and challenges. *Cancer Treat Rev*. 2021;99:102263.
196. Shafiq-Ul-Hassan M, Latifi K, Zhang G, Ullah G, Gillies R, Moros E. Voxel size and gray level normalization of CT radiomic features in lung cancer. *Sci Rep*. 2018;8(1):10545.

197. Park BW, Kim JK, Heo C, Park KJ. Reliability of CT radiomic features reflecting tumour heterogeneity according to image quality and image processing parameters. *Sci Rep*. 2020;10(1):3852.
198. de Sousa VML, Carvalho L. Heterogeneity in Lung Cancer. *Pathobiology*. 2018;85(1-2):96-107.
199. Chatterjee A, Vallières M, Forghani R, Seuntjens J. Investigating the impact of the CT Hounsfield unit range on radiomic feature stability using dual energy CT data. *Phys Med*. 2021;88:272-7.
200. Mali SA, Ibrahim A, Woodruff HC, Andrearczyk V, Müller H, Primakov S, et al. Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods. *J Pers Med*. 2021;11(9).
201. Pinna R, Campus G, Cumbo E, Mura I, Milia E. Xerostomia induced by radiotherapy: an overview of the physiopathology, clinical evidence, and management of the oral damage. *Ther Clin Risk Manag*. 2015;11:171-88.
202. Abdollahi H, Dehesh T, Abdolvand N, Rahmim A. Radiomics and dosiomics-based prediction of radiotherapy-induced xerostomia in head and neck cancer patients. *Int J Radiat Biol*. 2023:1-15.
203. Bang C, Bernard G, Le WT, Lalonde A, Kadoury S, Bahig H. Artificial intelligence to predict outcomes of head and neck radiotherapy. *Clin Transl Radiat Oncol*. 2023;39:100590.
204. Johansson AK, Omar R, Mastrovito B, Sannevik J, Carlsson GE, Johansson A. Prediction of xerostomia in a 75-year-old population: A 25-year longitudinal study. *J Dent*. 2022;118:104056.

205. Johansson AK, Johansson A, Unell L, Ekbäck G, Ordell S, Carlsson GE. Self-reported dry mouth in Swedish population samples aged 50, 65 and 75 years. *Gerodontology*. 2012;29(2):e107-15.
206. Rad M, Kakoie S, Niliye Brojeni F, Pourdamghan N. Effect of Long-term Smoking on Whole-mouth Salivary Flow Rate and Oral Health. *J Dent Res Dent Clin Dent Prospects*. 2010;4(4):110-4.
207. van Dijk LV, Noordzij W, Brouwer CL, Boellaard R, Burgerhof JGM, Langendijk JA, et al. (18)F-FDG PET image biomarkers improve prediction of late radiation-induced xerostomia. *Radiother Oncol*. 2018;126(1):89-95.
208. Beetz I, Schilstra C, Burlage FR, Koken PW, Doornaert P, Bijl HP, et al. Development of NTCP models for head and neck cancer patients treated with three-dimensional conformal radiotherapy for xerostomia and sticky saliva: the role of dosimetric and clinical factors. *Radiother Oncol*. 2012;105(1):86-93.
209. Soares I, Dias J, Rocha H, Khouri L, Lopes M, Ferreira B. Predicting xerostomia after IMRT treatments: a data mining approach. *Health and Technology*. 2018;8(1-2):159-68.
210. Araújo ALD, Moraes MC, Pérez-de-Oliveira ME, Silva VMD, Saldivia-Siracusa C, Pedroso CM, et al. Machine learning for the prediction of toxicities from head and neck cancer treatment: A systematic review with meta-analysis. *Oral Oncol*. 2023;140:106386.
211. Koçak B, Durmaz E, Ateş E, Kılıçkesmez Ö. Radiomics with artificial intelligence: a practical guide for beginners. *Diagn Interv Radiol*. 2019;25(6):485-95.

212. Anne-Leen D, Machaba S, Alex M, Bart S, Laurence B, Mike S, et al. Principal component analysis of texture features derived from FDG PET images of melanoma lesions. *EJNMMI Phys.* 2022;9(1):64.
213. Beddok A, Orlhac F, Calugaru V, Champion L, Ala Eddine C, Nioche C, et al. [18F]-FDG PET and MRI radiomic signatures to predict the risk and the location of tumor recurrence after re-irradiation in head and neck cancer. *Eur J Nucl Med Mol Imaging.* 2023;50(2):559-71.
214. Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med.* 2011;18(10):1099-104.
215. Mansournia MA, Geroldinger A, Greenland S, Heinze G. Separation in Logistic Regression: Causes, Consequences, and Control. *Am J Epidemiol.* 2018;187(4):864-70.
216. Li M, Zhang J, Zha Y, Li Y, Hu B, Zheng S, et al. A prediction model for xerostomia in locoregionally advanced nasopharyngeal carcinoma patients receiving radical radiotherapy. *BMC Oral Health.* 2022;22(1):239.
217. Kocak B, Baessler B, Bakas S, Cuocolo R, Fedorov A, Maier-Hein L, et al. CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. *Insights Imaging.* 2023;14(1):75.
218. Jiang W, Lakshminarayanan P, Hui X, Han P, Cheng Z, Bowers M, et al. Machine Learning Methods Uncover Radiomorphologic Dose Patterns in Salivary Glands that Predict Xerostomia in Patients with Head and Neck Cancer. *Adv Radiat Oncol.* 2019;4(2):401-12.
219. Beetz I, Schilstra C, van der Schaaf A, van den Heuvel ER, Doornaert P, van Luijk P, et al. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy

for head and neck cancer: the role of dosimetric and clinical factors.
Radiother Oncol. 2012;105(1):101-6.

220. Villa A, Abati S. Risk factors and symptoms associated with xerostomia: a cross-sectional study. Aust Dent J. 2011;56(3):290-5.

221. Dhull AK, Atri R, Dhankhar R, Chauhan AK, Kaushal V. Major Risk Factors in Head and Neck Cancer: A Retrospective Analysis of 12-Year Experiences. World J Oncol. 2018;9(3):80-4.

222. Morita I, Morioka H, Abe Y, Nomura T, Nakashima S, Sugiura I, et al. Discordance between hyposalivation and xerostomia among community-dwelling older adults in Japan. PLoS One. 2023;18(3):e0282740.

223. Ritchie H, Roser MR. Smoking 2013 [cited 2023 November 6th]. Available from: <https://ourworldindata.org/smoking>.

224. Ritchie H, Roser M. Alcohol Consumption 2018 [cited 2023 November 6th]. Available from: <https://ourworldindata.org/alcohol-consumption>.

225. Chaturvedi AK, D'Souza G, Gillison ML, Katki HA. Burden of HPV-positive oropharynx cancers among ever and never smokers in the U.S. population. Oral Oncol. 2016;60:61-7.

226. Jensen SB, Pedersen AM, Vissink A, Andersen E, Brown CG, Davies AN, et al. A systematic review of salivary gland hypofunction and xerostomia induced by cancer therapies: prevalence, severity and impact on quality of life. Support Care Cancer. 2010;18(8):1039-60.

227. Parkar S, Sharma A. Validation of European Organization for Research and Treatment of Cancer Head and Neck Cancer Quality of Life

Questionnaire (EORTC QLQ-H&N35) Across Languages: A Systematic Review. *Indian J Otolaryngol Head Neck Surg.* 2022;74(Suppl 3):6100-7.

228. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Logistic regression. *Perspect Clin Res.* 2017;8(3):148-51.

229. Forde E, Leech M, Robert C, Marignol L. Spatial Resampling Combined with Intensity Discretization Improves Reproducibility of Radiomic Features of the Parotid Gland. *European Congress of Medical Physics; Dublin, Ireland* 2022.

230. van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. *BMC Med.* 2023;21(1):70.

231. Aringhieri G, Fanni SC, Febi M, Colligiani L, Cioni D, Neri E. The Role of Radiomics in Salivary Gland Imaging: A Systematic Review and Radiomics Quality Assessment. *Diagnostics (Basel).* 2022;12(12).

232. Bhide SA, Davies M, Burke K, McNair HA, Hansen V, Barbachano Y, et al. Weekly volume and dosimetric changes during chemoradiotherapy with intensity-modulated radiation therapy for head and neck cancer: a prospective observational study. *Int J Radiat Oncol Biol Phys.* 2010;76(5):1360-8.

233. Loo H, Fairfoul J, Chakrabarti A, Dean JC, Benson RJ, Jefferies SJ, et al. Tumour shrinkage and contour change during radiotherapy increase the dose to organs at risk but not the target volumes for head and neck cancer patients treated on the TomoTherapy HiArt™ system. *Clin Oncol (R Coll Radiol).* 2011;23(1):40-7.

234. Zhang Y, Lin C, Wu J, Jiang X, Lee SWY, Tam SY, et al. A longitudinal evaluation of early anatomical changes of parotid gland in intensity

modulated radiotherapy of nasopharyngeal carcinoma patients with parapharyngeal space involvement. *J Med Radiat Sci.* 2017;64(3):188-94.

235. Ghosh A, Gupta S, Johny D, Vidyadhar Bhosale V, Pal Singh Negi M. A Study to Assess the Dosimetric Impact of the Anatomical Changes Occurring in the Parotid Glands and Tumour Volume during Intensity Modulated Radiotherapy using Simultaneous Integrated Boost (IMRT-SIB) in Head and Neck Squamous Cell Cancers. *Cancer Med.* 2021;10(15):5175-90.

236. Fiorino C, Rizzo G, Scalco E, Broggi S, Belli ML, Dell'Oca I, et al. Density variation of parotid glands during IMRT for head-neck cancer: correlation with treatment and anatomical parameters. *Radiother Oncol.* 2012;104(2):224-9.

237. Konings AW, Coppes RP, Vissink A. On the mechanism of salivary gland radiosensitivity. *Int J Radiat Oncol Biol Phys.* 2005;62(4):1187-94.

238. Teshima K, Murakami R, Tomitaka E, Nomura T, Toya R, Hiraki A, et al. Radiation-induced parotid gland changes in oral cancer patients: correlation between parotid volume and saliva production. *Jpn J Clin Oncol.* 2010;40(1):42-6.

239. van Dijk LV, Langendijk JA, Zhai TT, Vedelaar TA, Noordzij W, Steenbakkers RJHM, et al. Delta-radiomics features during radiotherapy improve the prediction of late xerostomia. *Sci Rep.* 2019;9(1):12483.

240. Forde E, Leech M, Robert C, Herron E, Marignol L. Influence of inter-observer delineation variability on radiomic features of the parotid gland. *Phys Med.* 2021;82:240-8.

241. Drummond JR, Newton JP, Abel RW. Tomographic measurements of age changes in the human parotid gland. *Gerodontology*. 1995;12(1):26-30.
242. Berger T, Noble DJ, Shelley LEA, McMullan T, Bates A, Thomas S, et al. Predicting radiotherapy-induced xerostomia in head and neck cancer patients using day-to-day kinetics of radiomics features. *Phys Imaging Radiat Oncol*. 2022;24:95-101.
243. Olver I. Xerostomia: a common adverse effect of drugs and radiation. *Australian Prescriber*. 2006;29(108).
244. Wolff A, Joshi RK, Ekström J, Aframian D, Pedersen AM, Proctor G, et al. A Guide to Medications Inducing Salivary Gland Dysfunction, Xerostomia, and Subjective Sialorrhea: A Systematic Review Sponsored by the World Workshop on Oral Medicine VI. *Drugs R D*. 2017;17(1):1-28.
245. Gomes M, Gutacker N, Bojke C, Street A. Addressing Missing Data in Patient-Reported Outcome Measures (PROMS): Implications for the Use of PROMS for Comparing Provider Performance. *Health Econ*. 2016;25(5):515-28.
246. Huang EP, O'Connor JPB, McShane LM, Giger ML, Lambin P, Kinahan PE, et al. Criteria for the translation of radiomics into clinically useful tests. *Nat Rev Clin Oncol*. 2023;20(2):69-82.
247. Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. *Radiother Oncol*. 2016;121(2):169-79.
248. Kothari G, Woon B, Patrick CJ, Korte J, Wee L, Hanna GG, et al. The impact of inter-observer variation in delineation on robustness of

radiomics features in non-small cell lung cancer. *Sci Rep*. 2022;12(1):12822.

249. Lin YC, Lin G, Pandey S, Yeh CH, Wang JJ, Lin CY, et al. Fully automated segmentation and radiomics feature extraction of hypopharyngeal cancer on MRI using deep learning. *Eur Radiol*. 2023;33(9):6548-56.

250. Ferrante M, Rinaldi L, Botta F, Hu X, Dolp A, Minotti M, et al. Application of nnU-Net for Automatic Segmentation of Lung Lesions on CT Images and Its Implication for Radiomic Models. *J Clin Med*. 2022;11(24).

251. Dutta K, Roy S, Whitehead TD, Luo J, Jha AK, Li S, et al. Deep Learning Segmentation of Triple-Negative Breast Cancer (TNBC) Patient Derived Tumor Xenograft (PDX) and Sensitivity of Radiomic Pipeline to Tumor Probability Boundary. *Cancers (Basel)*. 2021;13(15).

252. Orlhac F, Eertink JJ, Cottreau AS, Zijlstra JM, Thieblemont C, Meignan MA, et al. A guide to ComBat harmonization of imaging biomarkers in multicenter studies. *J Nucl Med*. 2021.

253. Da-Ano R, Visvikis D, Hatt M. Harmonization strategies for multicenter radiomics investigations. *Phys Med Biol*. 2020;65(24):24TR02.

254. Bettinelli A, Marturano F, Sarnelli A, Bertoldo A, Paiusco M. The ImSURE phantoms: a digital dataset for radiomic software benchmarking and investigation. *Sci Data*. 2022;9(1):695.

255. Ogura I, Sasaki Y, Oda T, Sue M, Yamaguchi H, Kameta A, et al. Structural Variations in Parotid Glands Induced by Radiation Therapy in Patients with Oral Carcinoma Observed on Contrast-Enhanced Computed Tomography. *Pol J Radiol*. 2017;82:561-6.

256. Wu VW, Ying MT, Kwong DL. Evaluation of radiation-induced changes to parotid glands following conventional radiotherapy in patients with nasopharyngeal carcinoma. *Br J Radiol*. 2011;84(1005):843-9.
257. Sanguineti G, Ricchetti F, Thomas O, Wu B, McNutt T. Pattern and predictors of volumetric change of parotid glands during intensity modulated radiotherapy. *Br J Radiol*. 2013;86(1031):20130363.
258. Ricchetti F, Wu B, McNutt T, Wong J, Forastiere A, Marur S, et al. Volumetric change of selected organs at risk during IMRT for oropharyngeal cancer. *Int J Radiat Oncol Biol Phys*. 2011;80(1):161-8.
259. Fiorentino A, Caivano R, Metallo V, Chiumento C, Cozzolino M, Califano G, et al. Parotid gland volumetric changes during intensity-modulated radiotherapy in head and neck cancer. *Br J Radiol*. 2012;85(1018):1415-9.
260. Ren G, Xu SP, Du L, Feng LC, Qu BL, Liu HX, et al. Actual anatomical and dosimetric changes of parotid glands in nasopharyngeal carcinoma patients during intensity modulated radiation therapy. *Biomed Res Int*. 2015;2015:670327.
261. Sato T, Ishii H. Regulation of hemodynamics in major salivary glands by parasympathetic vasodilation. *Journal of Oral Biosciences*. 2017;59:6.
262. Dirix P, Nuyts S, Van den Bogaert W. Radiation-induced xerostomia in patients with head and neck cancer: a literature review. *Cancer*. 2006;107(11):2525-34.
263. Bruvo M, Mahmood F. Apparent diffusion coefficient measurement of the parotid gland parenchyma. *Quant Imaging Med Surg*. 2021;11(8):3812-29.

264. Robar JL, Day A, Clancey J, Kelly R, Yewondwossen M, Hollenhorst H, et al. Spatial and dosimetric variability of organs at risk in head-and-neck intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys*. 2007;68(4):1121-30.

265. Morgan HE, Sher DJ. Adaptive radiotherapy for head and neck cancer. *Cancers Head Neck*. 2020;5:1.

266. Kreps S, Berges O, Belin L, Zefkili S, Petras S, Giraud P. Salivary gland-sparing helical tomotherapy for head and neck cancer: Preserved salivary function on quantitative salivary gland scintigraphy after tomotherapy. *Eur Ann Otorhinolaryngol Head Neck Dis*. 2016;133(4):257-62.

267. Brouwer CL, Steenbakkers RJ, Langendijk JA, Sijtsema NM. Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? *Radiother Oncol*. 2015;115(3):285-94.

268. Grégoire V, Jeraj R, Lee JA, O'Sullivan B. Radiotherapy for head and neck tumours in 2012 and beyond: conformal, tailored, and adaptive? *Lancet Oncol*. 2012;13(7):e292-300.

269. Castelli J, Simon A, Louvel G, Henry O, Chajon E, Nassef M, et al. Impact of head and neck cancer adaptive radiotherapy to spare the parotid glands and decrease the risk of xerostomia. *Radiat Oncol*. 2015;10:6.

270. Schwartz DL, Dong L. Adaptive radiation therapy for head and neck cancer-can an old goal evolve into a new standard? *J Oncol*. 2011;2011.

271. Castadot P, Lee JA, Geets X, Grégoire V. Adaptive radiotherapy of head and neck cancer. *Semin Radiat Oncol*. 2010;20(2):84-93.
272. Hamming-Vrieze O, van Kranen SR, Heemsbergen WD, Lange CAH, van den Brekel MWM, Verheij M, et al. Analysis of GTV reduction during radiotherapy for oropharyngeal cancer: Implications for adaptive radiotherapy. *Radiother Oncol*. 2017;122(2):224-8.
273. Brown E, Owen R, Harden F, Mengersen K, Oestreich K, Houghton W, et al. Predicting the need for adaptive radiotherapy in head and neck cancer. *Radiother Oncol*. 2015;116(1):57-63.
274. World Health Organisation (WHO). 6 Facts on Obesity 2024 [cited 2024 February 15th]. Available from: <https://www.who.int/news-room/facts-in-pictures/detail/6-facts-on-obesity>.
275. Gayar OH, Ruterbusch JJ, Elshaikh M, Cote M, Ghanem T, Hall F, et al. Oropharyngeal Carcinoma in Young Adults An Alarming National Trend. *Otolaryngology-Head and Neck Surgery*. 2014;150(4):594-601.
276. Wang YL, Zhu GP, Li DS, Wang Y, Wang ZY, Ji QH. Lymph node ratio as a survival risk factor for oropharyngeal squamous cell carcinoma patients with lymph node metastasis: A SEER-based analysis. *Head and Neck Oncology*. 2013;5(3).
277. Han C, Chen YJ, Liu A, Schultheiss TE, Wong JY. Actual dose variation of parotid glands and spinal cord for nasopharyngeal cancer patients during radiotherapy. *Int J Radiat Oncol Biol Phys*. 2008;70(4):1256-62.
278. Lee C, Langen KM, Lu W, Haimerl J, Schnarr E, Ruchala KJ, et al. Assessment of parotid gland dose changes during head and neck cancer radiotherapy using daily megavoltage computed tomography and

deformable image registration. *Int J Radiat Oncol Biol Phys*. 2008;71(5):1563-71.

279. Ger RB, Wei L, Naqa IE, Wang J. The Promise and Future of Radiomics for Personalized Radiotherapy Dosing and Adaptation. *Semin Radiat Oncol*. 2023;33(3):252-61.

280. Yu TT, Lam SK, To LH, Tse KY, Cheng NY, Fan YN, et al. Pretreatment Prediction of Adaptive Radiation Therapy Eligibility Using MRI-Based Radiomics for Advanced Nasopharyngeal Carcinoma Patients. *Front Oncol*. 2019;9:1050.

281. Lam SK, Zhang Y, Zhang J, Li B, Sun JC, Liu CY, et al. Multi-Organ Omics-Based Prediction for Adaptive Radiation Therapy Eligibility in Nasopharyngeal Carcinoma Patients Undergoing Concurrent Chemoradiotherapy. *Front Oncol*. 2021;11:792024.

282. Tanaka S, Kadoya N, Sugai Y, Umeda M, Ishizawa M, Katsuta Y, et al. A deep learning-based radiomics approach to predict head and neck tumor regression for adaptive radiotherapy. *Sci Rep*. 2022;12(1):8899.

283. Bertholet J, Anastasi G, Noble D, Bel A, van Leeuwen R, Roggen T, et al. Patterns of practice for adaptive and real-time radiation therapy (POP-ART RT) part II: Offline and online plan adaption for interfractional changes. *Radiother Oncol*. 2020;153:88-96.

284. Vásquez Osorio EM, Hoogeman MS, Al-Mamgani A, Teguh DN, Levendag PC, Heijmen BJ. Local anatomic changes in parotid and submandibular glands during radiotherapy for oropharynx cancer and correlation with dose, studied in detail with nonrigid registration. *Int J Radiat Oncol Biol Phys*. 2008;70(3):875-82.

285. Brock KK. Adaptive Radiotherapy: Moving Into the Future. *Semin Radiat Oncol.* 2019;29(3):181-4.
286. Green OL, Henke LE, Hugo GD. Practical Clinical Workflows for Online and Offline Adaptive Radiation Therapy. *Semin Radiat Oncol.* 2019;29(3):219-27.
287. van Beek S, Jonker M, Hamming-Vrieze O, Al-Mamgani A, Navran A, Remeijer P, et al. Protocolised way to cope with anatomical changes in head & neck cancer during the course of radiotherapy. *Tech Innov Patient Support Radiat Oncol.* 2019;12:34-40.
288. Ge X, Liao Z, Yuan J, Mao D, Li Y, Yu E, et al. Radiotherapy-related quality of life in patients with head and neck cancers: a meta-analysis. *Support Care Cancer.* 2020;28(6):2701-12.
289. Voordeckers M, Everaert H, Tournel K, Verellen D, Baron I, Van Esch G, et al. Longitudinal assessment of parotid function in patients receiving tomotherapy for head-and-neck cancer. *Strahlenther Onkol.* 2008;184(8):400-5.
290. Jiang N, Zhao Y, Jansson H, Chen X, Mårtensson J. Experiences of xerostomia after radiotherapy in patients with head and neck cancer: A qualitative study. *J Clin Nurs.* 2018;27(1-2):e100-e8.
291. Maes A, Huygh I, Weltens C, Vandeveld G, Delaere P, Evers G, et al. De Gustibus: time scale of loss and recovery of tastes caused by radiotherapy. *Radiother Oncol.* 2002;63(2):195-201.
292. Zeidler J, Kutschan S, Dörfler J, Büntzel J, Huebner J. Impact of nutrition counseling on nutrition status in patients with head and neck cancer undergoing radio- or radiochemotherapy: a systematic review. *Eur Arch Otorhinolaryngol.* 2024.

293. Rosi-Schumacher M, Patel S, Phan C, Goyal N. Understanding Financial Toxicity in Patients with Head and Neck Cancer: A Systematic Review. *Clin Med Insights Oncol.* 2023;17:11795549221147730.
294. Khan MN, Hueniken K, Manojlovic-Kolarski M, Eng L, Mirshams M, Khan K, et al. Out-of-pocket costs associated with head and neck cancer treatment. *Cancer Rep (Hoboken).* 2022;5(7):e1528.
295. Alhejoury HA, Mogharbel LF, Al-Qadhi MA, Shamlan SS, Alturki AF, Babatin WM, et al. Artificial Saliva for Therapeutic Management of Xerostomia: A Narrative Review. *J Pharm Bioallied Sci.* 2021;13(Suppl 2):S903-S7.
296. Kalsi H, McCaul LK, Rodriguez JM. The role of primary dental care practitioners in the long-term management of patients treated for head and neck cancer. *Br Dent J.* 2022;233(9):765-8.
297. Nonzee NJ, Dandade NA, Patel U, Markossian T, Agulnik M, Argiris A, et al. Evaluating the supportive care costs of severe radiochemotherapy-induced mucositis and pharyngitis : results from a Northwestern University Costs of Cancer Program pilot study with head and neck and nonsmall cell lung cancer patients who received care at a county hospital, a Veterans Administration hospital, or a comprehensive cancer care center. *Cancer.* 2008;113(6):1446-52.
298. D'Rummo KA, Miller L, TenNapel MJ, Shen X. Assessing the Financial Toxicity of Radiation Oncology Patients Using the Validated Comprehensive Score for Financial Toxicity as a Patient-Reported Outcome. *Pract Radiat Oncol.* 2020;10(5):e322-e9.

299. Verdonck-de Leeuw IM, van Bleek WJ, Leemans CR, de Bree R. Employment and return to work in head and neck cancer survivors. *Oral Oncol.* 2010;46(1):56-60.
300. Smith GL, Banegas MP, Acquati C, Chang S, Chino F, Conti RM, et al. Navigating financial toxicity in patients with cancer: A multidisciplinary management approach. *CA Cancer J Clin.* 2022;72(5):437-53.
301. Longo CJ, Gordon LG, Nund RL, Hart NH, Teleni L, Thamm C, et al. Clinical Management of Financial Toxicity-Identifying Opportunities through Experiential Insights of Cancer Survivors, Caregivers, and Social Workers. *Curr Oncol.* 2022;29(10):7705-17.
302. Wright P, Smith A, Roberts K, Selby P, Velikova G. Screening for social difficulties in cancer patients: clinical utility of the Social Difficulties Inventory. *Br J Cancer.* 2007;97(8):1063-70.
303. Elting LS, Chang YC. Costs of Oral Complications of Cancer Therapies: Estimates and a Blueprint for Future Study. *J Natl Cancer Inst Monogr.* 2019;2019(53).
304. Yong JH, Beca J, O'Sullivan B, Huang SH, McGowan T, Warde P, et al. Cost-effectiveness of intensity-modulated radiotherapy in oropharyngeal cancer. *Clin Oncol (R Coll Radiol).* 2012;24(7):532-8.
305. Khamdi S, Matangkasombut O, Lam-Ubol A. Non-pharmacologic interventions for management of radiation-induced dry mouth: A systematic review. *Oral Dis.* 2023.
306. Mohan V, Bruin NM, Steenbakkers RJHM, Noordzij W, Terhaard CHJ, de Keizer B, et al. Dose response modelling of secretory cell loss in salivary glands using PSMA PET. *Radiother Oncol.* 2022;177:164-71.

307. Cooper JS, Fu K, Marks J, Silverman S. Late effects of radiation therapy in the head and neck region. *Int J Radiat Oncol Biol Phys*. 1995;31(5):1141-64.
308. Eisbruch A. Reducing xerostomia by IMRT: what may, and may not, be achieved. *J Clin Oncol*. 2007;25(31):4863-4.
309. Berger T, Noble DJ, Yang Z, Shelley LEA, McMullan T, Bates A, et al. Assessing the generalisability of radiomics features previously identified as predictive of radiation-induced sticky saliva and xerostomia. *Phys Imaging Radiat Oncol*. 2023;25:100404.
310. Li Y, Sijtsema NM, de Vette SPM, Steenbakkers RJHM, Zhang F, Noordzij W, et al. Validation of the. *Radiother Oncol*. 2023;180:109458.
311. Fanizzi A, Scognamillo G, Nestola A, Bambace S, Bove S, Comes MC, et al. Transfer learning approach based on computed tomography images for predicting late xerostomia after radiotherapy in patients with oropharyngeal cancer. *Front Med (Lausanne)*. 2022;9:993395.
312. Zhou L, Zheng W, Huang S, Yang X. Integrated radiomics, dose-volume histogram criteria and clinical features for early prediction of saliva amount reduction after radiotherapy in nasopharyngeal cancer patients. *Discov Oncol*. 2022;13(1):145.
313. Jellema AP, Doornaert P, Slotman BJ, Leemans CR, Langendijk JA. Does radiation dose to the salivary glands and oral cavity predict patient-rated xerostomia and sticky saliva in head and neck cancer patients treated with curative radiotherapy? *Radiother Oncol*. 2005;77(2):164-71.
314. Spadarella G, Stanzione A, Akinci D'Antonoli T, Andreychenko A, Fanni SC, Ugga L, et al. Systematic review of the radiomics quality score

applications: an EuSoMII Radiomics Auditing Group Initiative. *Eur Radiol.* 2023;33(3):1884-94.

315. Garau N, Paganelli C, Summers P, Choi W, Alam S, Lu W, et al. External validation of radiomics-based predictive models in low-dose CT screening for early lung cancer diagnosis. *Med Phys.* 2020;47(9):4125-36.

316. Buvat I, Orlhac F. The Dark Side of Radiomics: On the Paramount Importance of Publishing Negative Results. *J Nucl Med.* 2019;60(11):1543-4.

317. Delgadillo R, Spieler BO, Deana AM, Ford JC, Kwon D, Yang F, et al. Cone-beam CT delta-radiomics to predict genitourinary toxicities and international prostate symptom of prostate cancer patients: a pilot study. *Sci Rep.* 2022;12(1):20136.

318. Yang P, Shan J, Ge X, Zhou Q, Ding M, Niu T, et al. Prediction of SBRT response in liver cancer by combining original and delta cone-beam CT radiomics: a pilot study. *Phys Eng Sci Med.* 2024.

319. Sellami S, Bourbonne V, Hatt M, Tixier F, Bouzid D, Lucia F, et al. Predicting response to radiotherapy of head and neck squamous cell carcinoma using radiomics from cone-beam CT images. *Acta Oncol.* 2022;61(1):73-80.

320. ClinicalTrials.gov Available from: <https://clinicaltrials.gov/search?term=radiomics&viewType=Table> [Date cited: February 16th 2024]

321. Nie K, Xiao Y. Radiomics in clinical trials: perspectives on standardization. *Phys Med Biol.* 2022;68(1).