

THE UNIVERSITY OF LIVERPOOL

PHD THESIS

---

**The Development of Machine  
Learning Methods for Head and  
Neck Cancer Prognosis**

---

*Author:*

Conor Whitley

*Supervisors:*

Dr. David Martin

Dr. Steve Barrett

Professor Marta Garcia-Finana

Dr Ruwanthi

Kolamunnage-Dona



## *Abstract*



# Contents

<b>Abstract</b>	<b>iii</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>1 Prognosis</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Materials and Methods . . . . .	3
1.3 Results . . . . .	9
1.4 Discussion . . . . .	19
1.5 Conclusion . . . . .	22



# List of Figures

1.1	Annotation of OSCC-containing areas in FTIR images. [A,D,G]: H&E image of a tissue core; [B,E,H]: FTIR image at $1650\text{cm}^{-1}$ ; [C,F,I]: Areas from which FTIR data was extracted for analysis . . .	5
1.2	Stratification of patients into high and low risk. A: Maximum log-rank statistic vs GA generation, plateauing around 40. B: whisker box plots of survival duration in each risk group. C: Kaplan-Meier Survival curves showing optimal risk stratification of the patient cohort, also shown are the log-rank statistic and corresponding p-value for the optimal groupings. . . . .	9
1.3	Median ROC and PR curves shown in solid lines; dashed lines represent baselines scores associated with random chance. AU-ROC and AUPRC scores are shown for each set of prognostic indicators. . . . .	11
1.4	Whisker boxplots of classification statistics calculated across all data subsets. AUROC (A); AUPRC (B); F1 (C); MCC (D); specificity (E); sensitivity (F); PPV (G); NPV (H). Boxes show the median, $25^{th}$ , and $75^{th}$ percentiles; whiskers extend to points that lie within 1.5 inter quartile ranges of the lower and upper quartiles; points lying outside this range are shown as individual diamonds. . . . .	12
1.5	Median ROC and PR curves shown in solid lines; dashed lines represent baselines scores associated with random chance. AU-ROC and AUPRC scores are shown for each set of prognostic indicators. . . . .	13

- 1.6 Whisker boxplots of classification statistics calculated across all data subsets. AUROC (A); AUPRC (B); F1 (C); MCC (D); specificity (E); sensitivity (F); PPV (G); NPV (H). Boxes show the median, 25<sup>th</sup>, and 75<sup>th</sup> percentiles; whiskers extend to points that lie within 1.5 inter quartile ranges of the lower and upper quartiles; points lying outside this range are shown as individual diamonds. 14
- 1.7 Kaplan-Meier survival curves for each risk group according input variables. Low-risk:blue, high-risk:red. Confidence intervals are computed using the exponential Greenwood method [?]. . . . . 16
- 1.8 Cox proportional hazards model patient simulation . . . . . 18



# List of Tables

1.1	Characteristics of the sample cohort . . . . .	10
1.2	Median classification statistics . . . . .	15
1.3	Cox proportional hazards model fit statistics . . . . .	17



# 1 Prognosis

## 1.1 Introduction

Head and neck squamous cell carcinomas (HNSCCs) are the sixth most common form of cancer worldwide, amounting to over 500,000 new cases annually. The majority of head and neck cancers are squamous cell carcinomas which originate in the upper aerodigestive epithelium. The development of oropharyngeal squamous cell carcinoma (OPSCC) is known to be linked to carcinogenic human papillomavirus (HPV), accounting for approximately 51.8% - 71% of cases [1].

Oral squamous cell carcinoma (OSCC) is the 8th most common form of cancer in the UK [2] with a recent increase in incidence reported [3, 4]. In contrast, OSCCs are rarely mediated by HPV [5] and the majority of cases are typically associated with exposure to carcinogens present in tobacco and alcohol [1, 3, 6].

The regions in which head and neck tumours typically develop are anatomically complex and play a vital physiological role in the patient; early diagnosis and selection of appropriate treatment should aim to increase patient survival while maximising retention of vital organ function.

A key issue facing clinical decision makers is the determination of the optimal course of treatment for a patient. In cases where lower biological aggression can be demonstrated, a de-escalation of therapy may be possible [7, 8].

Conversely, early identification of cases with poor prognosis could allow targeting for neo-adjuvant therapy. However, identification of these cases prior to surgery is not insignificant, even given the acknowledged prognostic importance of extra-nodal extension (ENE) [9]. Physical methods such as magnetic resonance imaging (MRI) and computed tomography (CT) have proven to be of limited benefit [10] despite improvements in analysis techniques utilising deep learning models [11].

Previous studies [12, 13, 14, 15] have hypothesised that tumours with ENE which may be responsive to novel therapeutic treatment may carry a distinct molecular fingerprint; the identification of which may allow for screening of patients towards appropriate treatment. Despite small studies identifying putative biomarkers of ENE that might be of use in the analysis of diagnostic biopsy material [16], none of the proposed biomarkers or molecular fingerprints of aggression has progressed into practice. For approximately 50% of HPV negative HNSCC patients, current treatment plans are ineffective [17, 18]. Neo-adjuvant therapy has the potential to improve prognoses, and aid clinical decision making if applicable cases can be determined at the time of diagnosis.

### infrared

Fourier transform infra-red (FTIR) microscopy is a well established technique, which has been utilised in a range of biomedical applications in recent years. Due to its ability to access chemical information present within the sample; FTIR microscopy data and accompanied multivariate analysis has been used to diagnose cancer in biofluids [19, 20, 21], surgically resected tissue [22, 23], and cells [24, 25, 26]. FTIR microscopy allows for imaging of sample specimens at thousands of infra-red wavelengths simultaneously using a typical spectrometer. This is achieved through the use of a broadband light source and Michelson interferometer set up. Individual chemical spectra are then obtained by performing a Fourier transform on the resulting interferogram. Marginal spectral differences in biochemical compounds of interest are typically located in a

region known as the fingerprint region ( $1000\text{cm}^{-1}$ - $1800\text{cm}^{-1}$ ). It is differences in these absorption bands which contain information which can be utilised to discriminate between samples of interest.

This work explores the potential efficacy of FTIR microscopy in combination with a known prognostic biomarker:  $\alpha$ -smooth muscle actin (ASMA) expression, as a method of identifying ENE prior to surgery. Previous work [27, 28, 29] has explored the efficacy of ASMA and SERPINE1 [27] as predictive variables for extra capsular spread (ECS), and as prognostic biomarkers for OSCC. ASMA expression is closely associated with the presence of activated fibroblasts, also known as myofibroblasts in tumour associated stroma. The degree of ASMA expression can be interrogated through the use of appropriate chemical stains, and evaluated using an optical microscope.

## 1.2 Materials and Methods

### Tissue preparation

The dataset comprised FTIR spectra taken from primary tumour sites of 29 patients with a diagnosis of OSCC. The specimens are a subset of those arranged in a previously described tissue microarray (TMA; [27]). Inclusion criteria for this study were: a diagnosis of OSCC; the presence of OSCC in the TMA core; the ability to co-register adjacent H&E stained and FTIR imaged sections; a follow-up period after surgery of at least 24 months; HPV negative. Patients gave written, informed consent and the study was undertaken under ethical approval (Northwest - Liverpool Central REC number EC47.01). All samples were 1mm diameter cores of formalin fixed paraffin embedded (FFPE) tissue arranged into a tissue micro array.

Four adjacent sections of  $4\mu\text{m}$  thickness were taken from the TMA, the first and last sections were stained with haematoxylin and eosin (H&E) and used to assess the presence and location of tumour material in the 2nd and 3rd sections. Specimens were removed from the sample set if no clear area containing predominantly tumour cells was discernable in the H&E stained sections. Samples were also removed from the sample set if the outline of the regions containing tumour cells were markedly different between the 1st and 4th sections. Images of stained sections were scanned using an Aperio CS2 scanner (Leica Biosystems) and used for IR image annotation. The 2nd and 3rd TMA sections were mounted onto  $\text{CaF}_2$  disks for FTIR microspectroscopy.

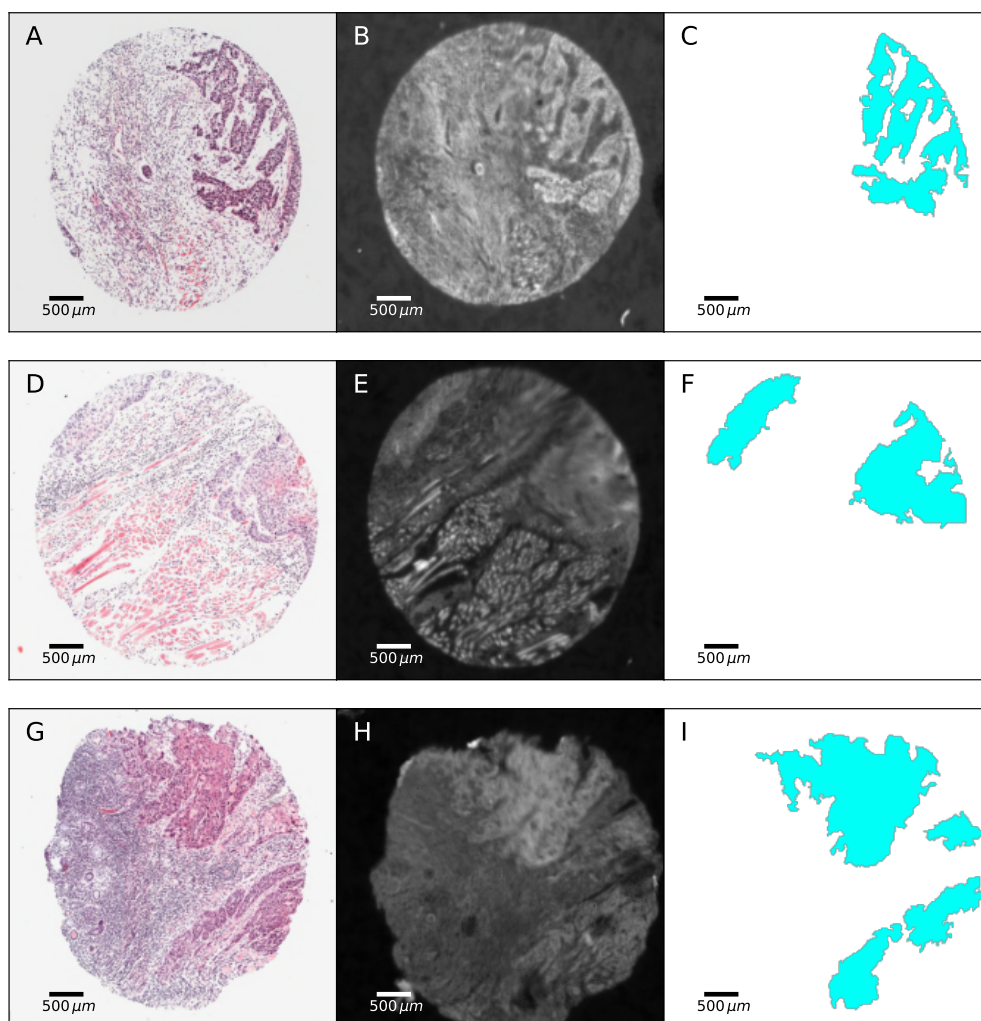


FIGURE 1.1: Annotation of OSCC-containing areas in FTIR images. [A,D,G]: H&E image of a tissue core; [B,E,H]: FTIR image at  $1650\text{cm}^{-1}$ ; [C,F,I]: Areas from which FTIR data was extracted for analysis

Figure 1.1 shows three examples of tumour areas selected from H&E sections and used for further analysis. Selections were annotated using GIMP [30] image manipulation software onto IR images at  $1650\text{cm}^{-1}$  to ensure spectra were extracted accurately. The obtained mask was then used to compile spectra of tumour regions indexed by patient identity and corresponding metadata.

## FTIR Microspectroscopy

FTIR measurements of TMA cores were taken at room temperature using a Varian Cary 670-FTIR spectrometer with an attached Varian Cary 620-FTIR microscope produced by Varian (now Agilent Technologies, Santa Clara CA, USA); with a liquid nitrogen cooled  $128 \times 128$  pixel mercury-cadmium-telluride (MCT) focal plane array with an effective field of view for each pixel of  $5.5 \mu\text{m}$ . The sample stage was enclosed in a perspex box and pumped with dry air until a humidity of 1% was achieved in order to mitigate the effects of water contributions on measured IR spectra. Images were acquired at a resolution of  $6 \text{cm}^{-1}$  over a spectral range of  $990 \text{cm}^{-1}$  to  $3800 \text{cm}^{-1}$  using a co-addition of 128 scans. Attenuator and integration time of the focal plane array (FPA) were chosen to gain the maximum signal-to-noise ratio. Background scans were acquired using a blank  $\text{CaF}_2$  disk situated within the perspex box before each session of measurements.

## Data Preprocessing & Analysis

Selection of tissue areas to include in the analysis was undertaken by a consultant oral pathologist (AT), who identified regions containing high proportions of tumour cells on the H&E images. These were subsequently co-registered with IR images at  $1650 \text{cm}^{-1}$  (the amide-I peak), from the same tissue core in order to extract IR data for analysis.

In order to correct for atmospheric scattering, extracted spectra were pre-processed using an open-source extended multiplicative scattering correction (EMSC) code provided by Kohler et.al [31]. An unsupervised quality control check of all data, to eliminate anomalous spectra through the use of the multivariate Hotelling's  $T^2$  statistic [32, 33]; spectra determined to have a  $T^2$  value lying outside the



95<sup>th</sup> percentile were deemed to be anomalous and were omitted from further analysis.

The following preprocessing steps were carried out on the dataset before a final classification step was performed using a logistic regression (LR) classifier. Vector normalisation was used in order to account for sample thickness; wavenumber absorbance features were mean-centered; and variance scaled to one; before a final principal component analysis (PCA) step to reduce dimensionality of the dataset. Seven principal components were taken to assist convergence when fitting the LR classifier. A large L2 regularisation term ( $1 \times 10^5$ ) was applied to the objective function when fitting the LR model in order to mitigate the potential for overfitting.

Scientific Python packages [34, 35, 36] were used to implement classification models and survival analysis. The classification power of an FTIR spectrum as a biomarker was estimated using the area under both the receiver operating characteristic (ROC) curve (AUROC), and the precision-recall (PR) curve (AUPRC). In order to obtain an estimate of the variability of the classification power of FTIR, bootstrap out-of-bag sampling was utilised as follows. A training data set was constructed by drawing 80% of patients in the total dataset without replacement. The remaining 20% was used as the "out of bag" test set on which fitted models were evaluated, and statistics calculated. This process was repeated 100 times ensuring that no two sample sets were identical. When fitting the LR model, data points were inversely weighted against differing number of acquired spectra per patient and by risk group to mitigate the imbalanced nature of the dataset. Predictions of risk group from the LR model are presented as a list of probabilities for each risk group. Final prediction scores for each patient are taken to be the median probability predicted for each patient.

ref to Matthews?

Calculated statistics included: AUROC, Matthew's correlation coefficient (MCC), specificity, sensitivity, positive predictive value (PPV), and negative predictive

value (NPV) — all statistics were calculated using appropriate weightings to compensate for class/patient imbalance and to ensure a classification statistics were not skewed. Prognostic efficacy was investigated using Kaplan-Meier survival-analysis, a Cox proportional hazards regression, and a log-rank test.

### **Prediction of patient outcomes using**

Rather than utilising arbitrary outcome cutoffs to identify degree of risk, the cohort was stratified into "high" and "low" risk categories. The choice of risk group for each patient was determined through an optimisation routine which maximised the log-rank statistic with respect to the groupings of patients solely using outcome data. The optimisation procedure was performed using a genetic algorithm (GA) based approach utilising the distributed evolutionary algorithms for python (DEAP) library [37]. The "individuals" involved in the GA routine are vectors comprising the identity of the risk group of each patient in the analysis. The "fitness" of an individual set is the log-rank statistic — calculated using the patient risk groups specified by that individual.

Maximum individual fitness of the generation against the generation number which showed a plateauing of the log-rank statistic at around 40 (fig. 1.2 A). The resulting risk groups show a clear distinction in clinical outcomes (fig. 1.2 B,C).

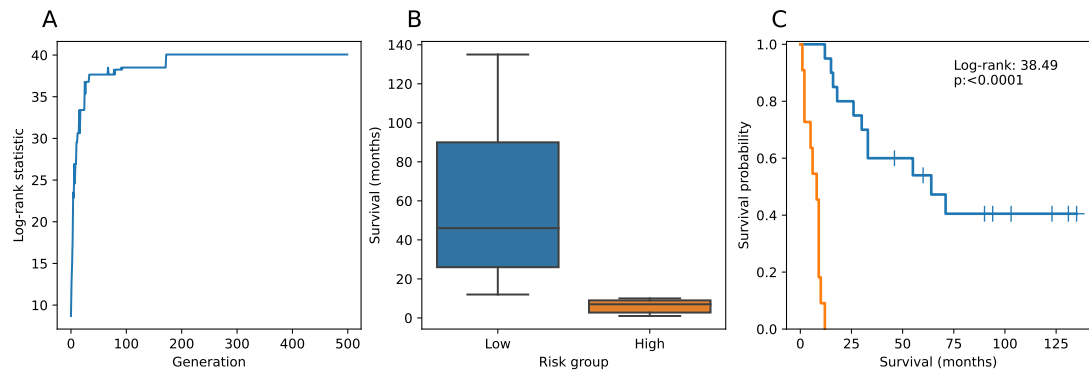


FIGURE 1.2: Stratification of patients into high and low risk. A: Maximum log-rank statistic vs GA generation, plateauing around 40. B: whisker box plots of survival duration in each risk group. C: Kaplan-Meier Survival curves showing optimal risk stratification of the patient cohort, also shown are the log-rank statistic and corresponding p-value for the optimal groupings.

The risk groups derived from the GA process were then utilised as prediction targets for classification. This process sought to identify objective groupings to identify using FTIR and ASMA as predictive variables.

## 1.3 Results

The inclusion criteria designated in materials and methods gave a sample set of 29 patients from the original 102 previously described [27]. Of these 29 patients, 19 remained alive 12 months after surgery, while 14 survived to two years. The cohort was representative of this original patient set except that the male:female ratio was inverted (table 1.1) but now corresponds more closely with larger datasets (table 1.1; final column). In agreement with the published cohort, the sample set was enriched for cases with ENE (i.e. poor prognosis) in comparison to the general HNSCC population [38].

TABLE 1.1: Characteristics of the sample cohort

	All (N=29)	Outcome at 12 months		Outcome at 24 months		Original cohort [27]  N=102	Larger, local cohort [38]  N=489
		Dead	Alive	Dead	Alive		
Age (years)							
Mean	60	70.4	56.3	64.5	57.2	60	
Range	29-85	59-85	29-72	48-85	29-68	29-89	
Median	61	70.5	56.5	64	60		
$\alpha$ -SMA							
High/Intermediate	26 (94)					60 (64)	ND
Low	3 (6)					33 (36)	ND
Gender							
F	7 (24) <sup>†</sup>	0	7	2	5	57 (56) <sup>a</sup>	187 (38) <sup>b</sup>
M	22 (76)	10	12	13	9	45 (45)	302 (62)
T Stage*							
1	1 (3)	0	1	0	1	8 (8)	123 (25) <sup>b</sup>
2	14 (48)	4	10	7	8	57 (54)	175 (35)
3	2 (7)	0	2	2	0	12 (11)	47 (10)
4	8 (28)	4	4	3	5	20 (19)	144 (30)
4a	4 (14)	2	2	2	2	9 (8)	
N Stage <sup>#</sup>							
0	7 (24)	2	4	1	5	38 (37) <sup>b</sup>	314 (64) <sup>c</sup>
1	7 (24)	2	4	2	4	18 (17)	64 (14)
2a	1 (3)	1	0	1	0		101(20)
2b	16 (55)	4	9	8	5	45 (44)	
2c	3 (10)	1	2	3	0		
Pathological Site							
Floor of mouth	8 (28)	2	6	4	4	35 (34) <sup>b</sup>	162 (33)
Other	12 (41)	5	7	5	7	24 (24)	183 (36)
Tongue	9 (31)	3	6	6	3	37 (36)	144 (30)

a:  $p < 0.005$ b:  $p = \text{NS}$ c:  $p < 0.00001$ 

\*: T stage 1+2 v 3+4

†: numbers in parenthesis are percentages

#: N stage 0 v 1 v 2

FTIR and ASMA data from the reduced sample set were evaluated as prognostic indicators of death within one year of surgery, both separately and together

(Figure 1.3). A total of 168,460 FTIR spectra were obtained from the 19 patients who survived beyond one year and 96,402 spectra were obtained from 10 patients who died within 12 months.

### Evaluation of individual spectra

Using predictions derived from individual spectra only, ROC and PR analysis were performed (fig. 1.3). The ASMA model shows poor performance across all thresholds for both analyses; the FTIR model is slightly better and is a reasonable predictor in its own right. However, the combined model shows superior ROC and PR curves overall. The baseline shown in fig. 1.3[B] corresponds to the prevalence of spectra labelled as high risk.

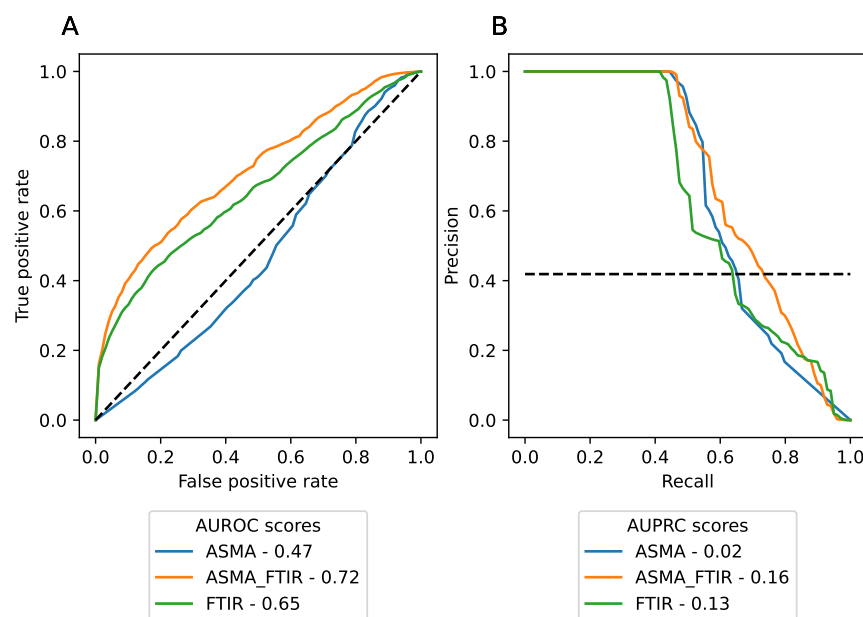


FIGURE 1.3: Median ROC and PR curves shown in solid lines; dashed lines represent baseline scores associated with random chance. AUROC and AUPRC scores are shown for each set of prognostic indicators.

Figure 1.4 shows distributions of classification statistics calculated for each set of prognostic variables. As shown in fig. 1.3[A,B] AUROC and AUPRC scores of the FTIR models show promising scores, but also a small spread in comparison

to the ASMA model. Many other statistics show the combined model as the most effective, with the exception of the sensitivity score.

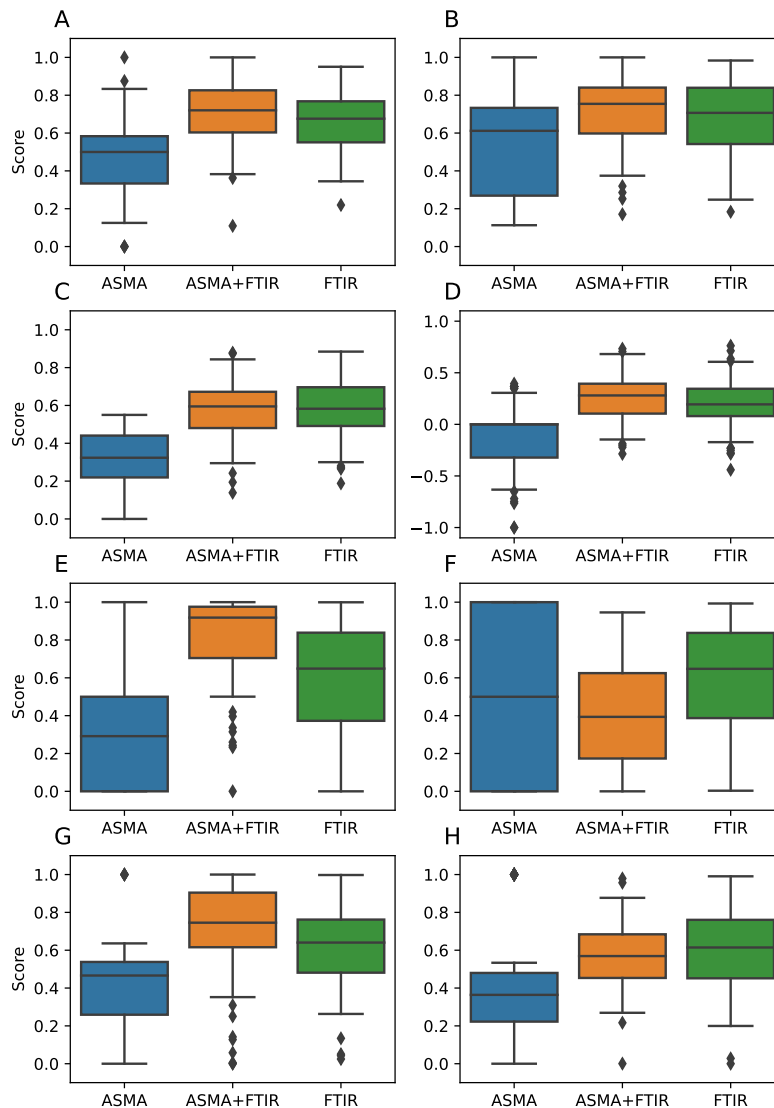


FIGURE 1.4: Whisker boxplots of classification statistics calculated across all data subsets. AUROC (A); AUPRC (B); F1 (C); MCC (D); specificity (E); sensitivity (F); PPV (G); NPV (H). Boxes show the median, 25<sup>th</sup>, and 75<sup>th</sup> percentiles; whiskers extend to points that lie within 1.5 inter quartile ranges of the lower and upper quartiles; points lying outside this range are shown as individual diamonds.

### Evaluation on a patient basis

To obtain a prediction of risk group for each patient, final prediction scores were taken as the median probability predicted across all spectra for any given

patient.

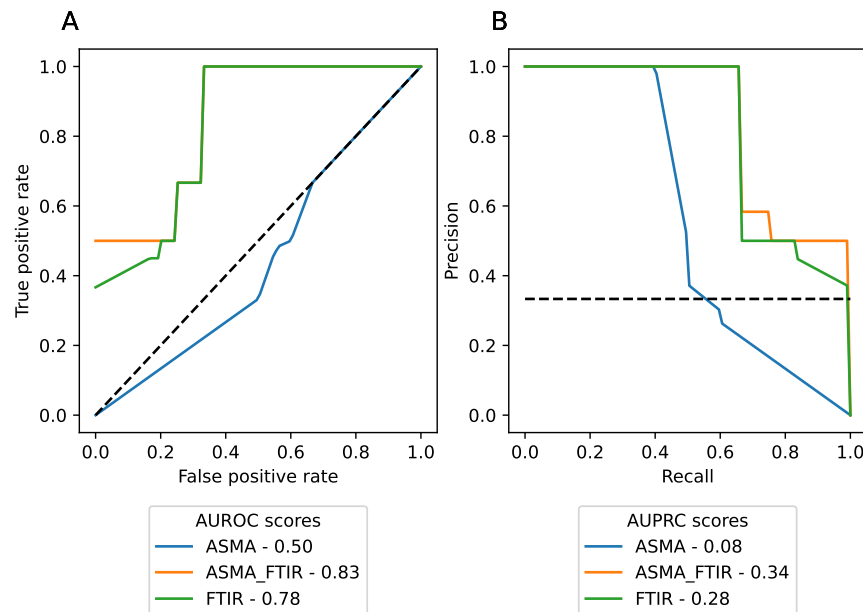


FIGURE 1.5: Median ROC and PR curves shown in solid lines; dashed lines represent baseline scores associated with random chance. AUROC and AUPRC scores are shown for each set of prognostic indicators.

The median AUROC obtained from FTIR alone was 0.89 (Figure 1.6 [a,c]); incorporation of the ASMA data into this analysis, increased the AUROC slightly to 0.92; while ASMA alone achieved a significantly poorer score of 0.46. Precision and recall scores both remained high for the FTIR and combined ASMA/FTIR models across a range of decision thresholds — indicating that both models can balance both statistics effectively, and that imbalance in the dataset was not detrimental (Figure 1.6 [B]).

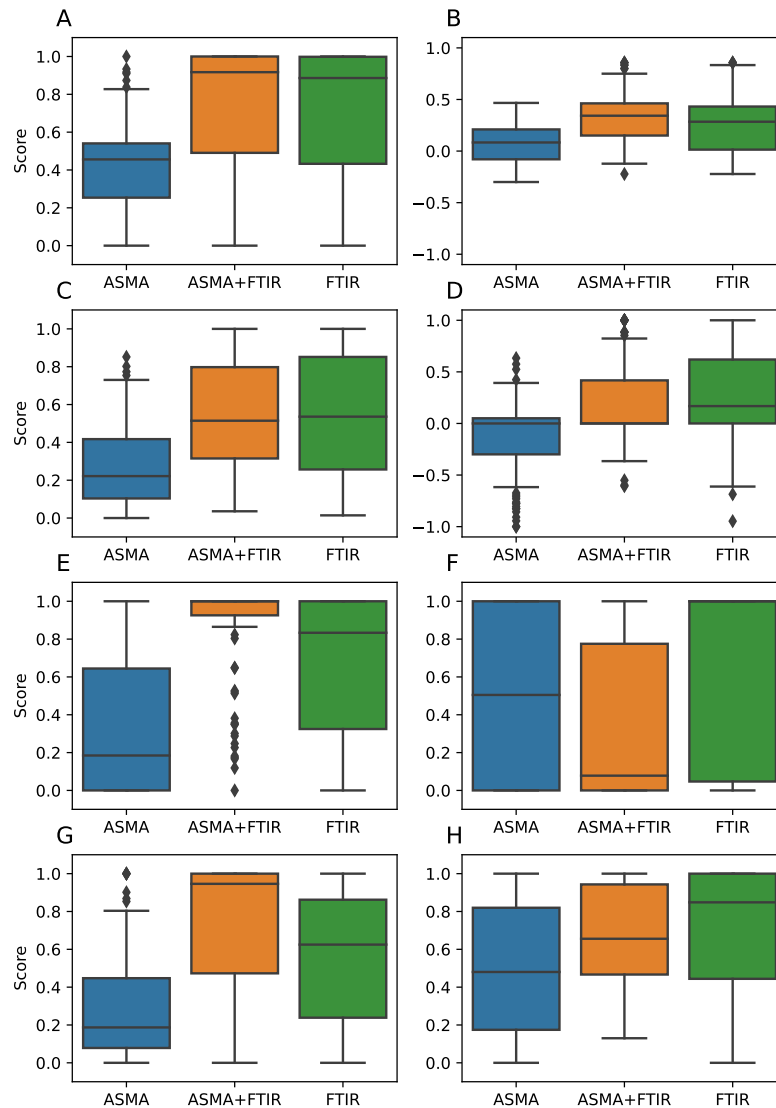


FIGURE 1.6: Whisker boxplots of classification statistics calculated across all data subsets. AUROC (A); AUPRC (B); F1 (C); MCC (D); specificity (E); sensitivity (F); PPV (G); NPV (H). Boxes show the median, 25<sup>th</sup>, and 75<sup>th</sup> percentiles; whiskers extend to points that lie within 1.5 inter quartile ranges of the lower and upper quartiles; points lying outside this range are shown as individual diamonds.

Furthermore, additional classification statistics produced comparable conclusions, showing that the combined model was a good predictor of poor outcome (Figure 1.6 [A-H]).

A comparison of the median scores of these statistics shows that ASMA alone



TABLE 1.2: Median classification statistics. Classification thresholds (table 1.2) used to convert probabilities to binary decisions were determined to be those that maximised the log-rank statistic.

Variables	AUROC	AUPRC	F1	MCC	Spec	Sens	PPV	NPV	Thresh
ASMA	0.46	0.41	0.22	0.00	0.18	0.50	0.19	0.48	0.48
ASMA+FTIR	0.92	0.85	0.51	0.00	1.00	0.08	0.95	0.66	0.69
FTIR	0.89	0.79	0.54	0.17	0.83	1.00	0.62	0.85	0.34

is a poor predictive variable for this dataset with low scores in all metrics (table 1.2), while the FTIR alone scores are consistently high. The combined model shows high specificity and low sensitivity, with a high PPV and moderate NPV indicating numerous false negatives.

The classification threshold for each model was used to assign patients to high or low risk groups and survival analyses were undertaken. As expected, the use of ASMA alone did not show good separation of the high and low risk groups as shown in Kaplan-Meier curves (Figure 1.7a) and, indeed, the risk group predictions were inverted to that expected. The use of FTR alone to predict risk showed significant separation of the groups (Figure 1.7c;  $p=0.01$ ), while the combined model produced good separation with a highly significant  $p$  value (Figure 1.7b;  $p<0.005$ ).

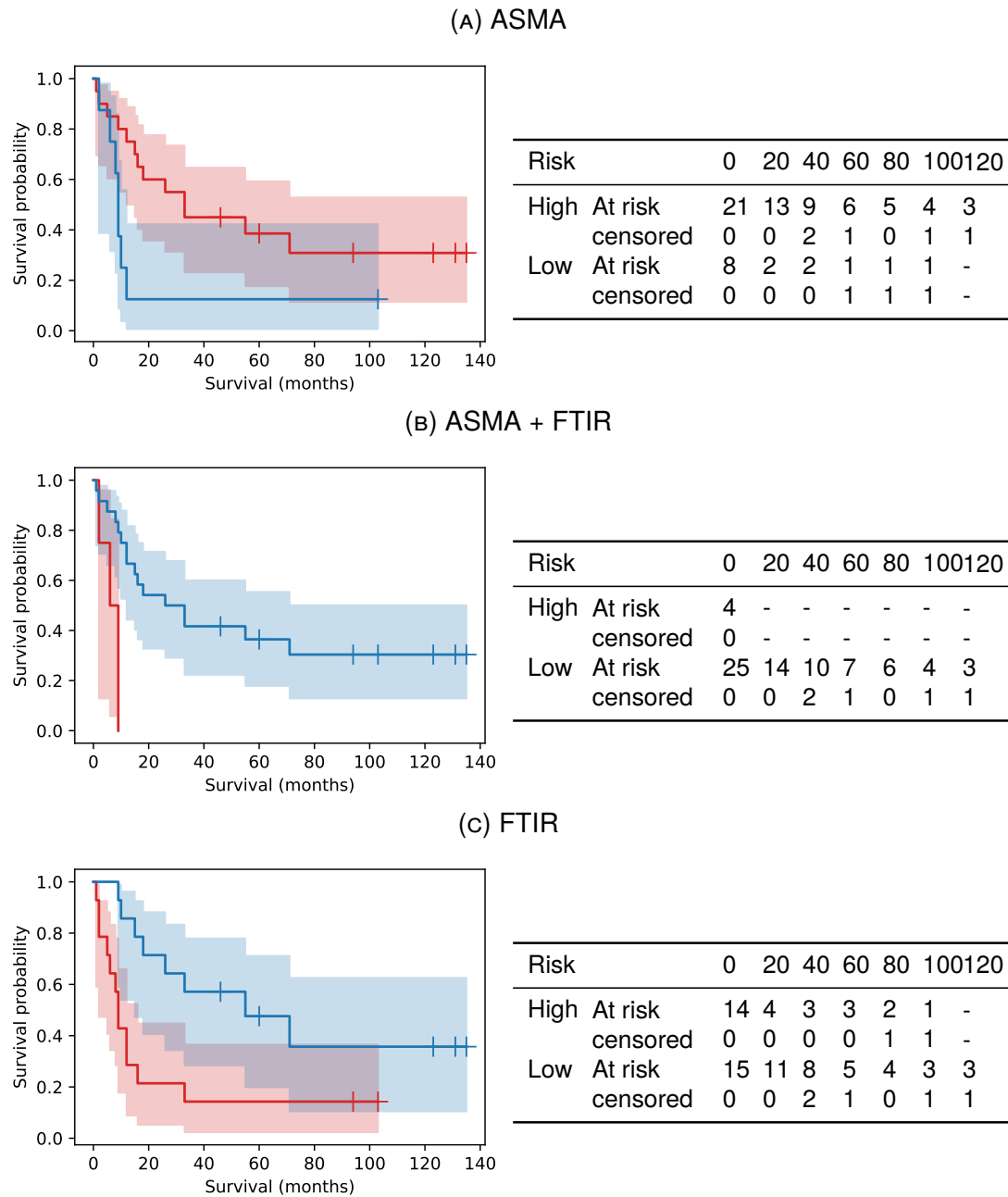


FIGURE 1.7: Kaplan-Meier survival curves for each risk group according input variables. Low-risk:blue, high-risk:red. Confidence intervals are computed using the exponential Greenwood method [39].

A univariate Cox proportional hazard model was fit to prediction scores obtained from the LR model to assess the prognostic utility of the prediction score before conversion to a binary decision. Both models using FTIR data have significantly higher hazard ratios than the pure ASMA model suggesting the LR classifier is

able to stratify risk groups effectively. However the 95% confidence intervals for all models spanned a wide range.

TABLE 1.3: Cox proportional hazards model fit statistics

Variables	Coefficients	z	p	Hazard ratio
ASMA	-0.03	-0.02	0.98	0.97 (0.06-15.08)
ASMA+FTIR	1.84	2.12	0.03	6.29 (1.14-34.59)
FTIR	2.02	2.07	0.04	7.52 (1.12-50.62)

Figures 1.8a to 1.8c show predicted survival curves for five simulated patients. Prediction scores for patients vary between 0 (definitely low risk) and 1 (definitely high risk) showing how LR predictions translate to expected survival outcomes. As expected fig. 1.8a showed very little stratification between simulated predictions; this is consistent with a hazard ratio of  $\sim 1$  as shown in table 1.3. The combined model shows a much better level of stratification by risk than ASMA alone with distinct survival curves clearly visible. The FTIR model also shows distinct survival curves consistent with a large hazard ratio. The addition of ASMA data appears to be detrimental to the overall prognostic utility of the combined model.

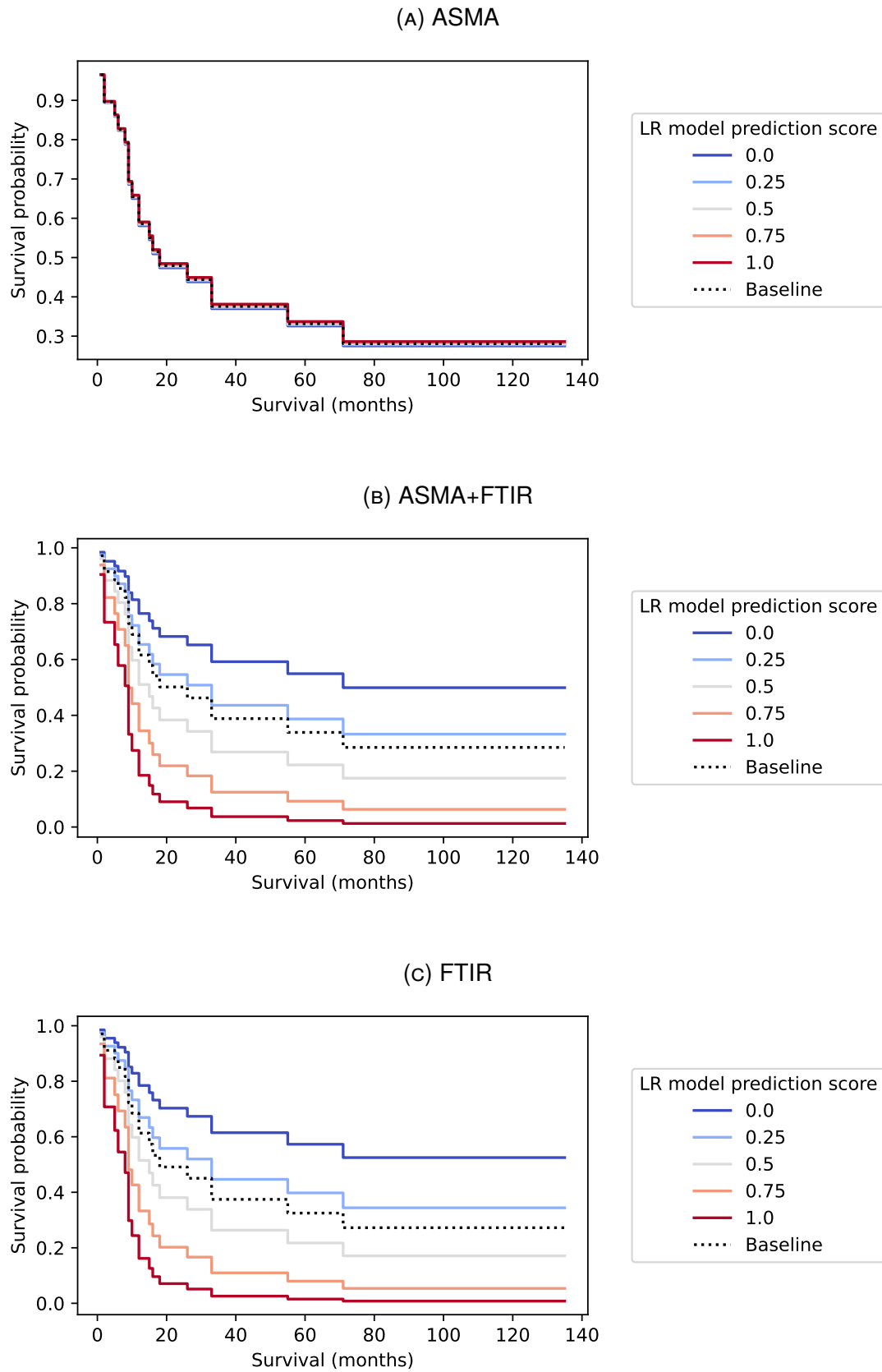


FIGURE 1.8: Predicted survival curves for five simulated patients varying between 0 and 1. The hazard baseline is represented by a dotted line.

## 1.4 Discussion

It has previously been established that the presence of ENE is a marker of poor prognosis in oral cancer. However, the presence of ENE cannot be firmly established until after pathological inspection of surgically removed nodal material. This prevents the use of neo-adjuvant treatment in those patients with poor prognosis, for whom current treatment regimens do not work. It has been shown that the chemical 'fingerprint' of the primary tumour, readily observed through the use of FTIR microscopy is able to identify individuals whose disease will progress following treatment.

Much of the existing literature concerning prognostic indicators utilises disease specific survival and overall survival as indicators of patient prognoses. Stratification of these measures into groups according to risk is typically decided according to a number of years e.g. one year, two years. The decision to use a cut off threshold of a number of years is arbitrary, thus, in order to determine a threshold decided objectively. A GA approach was used to stratify patients in this cohort into either a low or high risk group, where the log-rank statistic calculated from a log-rank test was maximised — indicating the best separation. This threshold was determined to be 11 months and was used as the *deen???*

Results presented here suggest that FTIR spectra can be used to stratify patients into useful clinical risk groups. Models utilising a combination of FTIR and ASMA data showed some diminished prognostic utility — suggesting that the addition of ASMA data was detrimental. Many other statistics show the combined model as the most effective, with the exception of the sensitivity score; indicating a high proportion of false negatives and thus suggesting that low risk patients are often misclassified and would be under treated in a clinical setting. The FTIR model scores are generally high, indicating that FTIR spectra can serve as a prognostically useful biomarker. A slightly lower PPV however

suggests the FTIR model incurs more false positives — indicating that patients with better prognoses may be given inappropriate treatment.

A significant improvement was observed when taking the median prediction score for each patient. This is potentially a reflection of the fact that the molecular fingerprint for poor prognosis varies in magnitude across the measured tumour section — genetic heterogeneity within OSCC lesions has been noted previously [40]. The aggregation of scores across an entire tumour section may cause a 'regularising' effect, thus mitigating the effects of overfitting to specific chemical fingerprints. This regularisation effect may be dependent upon the size of the measured tumour section present within the core. Data subsets containing patients with relatively few measured spectra may experience lower scores as a result; possibly explaining the large degree of variation within sample scores to some extent.

Survival analysis showed that groups allocated by the classifier had significantly different outcomes. The combined model (fig. 1.7b) is nearly able to perfectly replicate survival the ideal survival curves in (fig. 1.2c). Model predictions using FTIR data showed a marked improvement over the pure ASMA model. Hazard ratios of 6.29 and 7.5 for the ASMA+FTIR and FTIR model respectively show that the prognoses of patients allocated to the high-risk group are significantly poorer. 95% confidence intervals for hazard ratios span a large range (FTIR: 1.12-50.62; ASMA+FTIR: 1.14-34.59). This is indicative of a high degree of heterogeneity in the sample cohort and would justify further exploration with a larger cohort in the future.

The use of FTIR in clinical diagnostics is growing quickly, however relatively little work has been aimed towards prognostic biomarkers. The combination of both FTIR microscopy with techniques more familiar to oncologists such as immunohistochemical staining has the potential to improve prognostic predictive capabilities significantly as shown in this work. Previous studies [41, 42, 43, 29,

28, 27, 44] have investigated the viability of a range of prognostic biomarkers for head and neck cancer, with varying degrees of success. Many previously analysed biomarkers are measured on surgically resected tissue — limiting the potential for timely treatment. What is needed are prognostic biomarkers which can be measured in biopsy tissue prior to surgery. The discovery of effective prognostic biomarkers has been difficult, and thus far has largely focused on immunohistochemistry techniques. MRI has also been utilised [45, 46, 47] to measure physical attributes such as: tumour thickness, depth of invasion, and the presence of sub-volumes in a non-invasive manner. However, MRI based techniques often quantify biomarkers inaccurately when validated against direct measurements of pathological staging sections [48, 49].

Zawlik et.al [50] investigated FTIR coupled with PCA to investigate the efficacy of chemotherapy in triple-negative breast cancer. They determined that it was possible to monitor changes in the biochemical composition of the tissue in order to monitor the effectiveness of received treatment. Butler et.al [51] have undertaken development of a high-throughput ATR-FTIR based instrument for use in biofluid assays. Their work concluded that it was possible to triage brain cancer using utilising FTIR spectroscopy of biofluid samples. Their analysis comprised a large retrospective cohort of 724 patients with a range of brain cancer subtypes and stages. They utilised a binary support vector machine (SVM) classifier, and were able to achieve a sensitivity and specificity of 93.2% and 92.8% respectively.

Many other prognostic biomarkers exist [52, 53, 54] but a large proportion are still in the "discovery phase" — requiring further study to ascertain prognostic benefit [55]. The use of FTIR within clinical diagnostics likely fits into this category, as many potential barriers facing other potential biomarkers are still present.

A relatively small sample set was a key issue facing this study due to the difficulty in acquiring and imaging large numbers of samples. Despite attempts to determine the feasibility of FTIR as a prognostic tool through multiple sampling of the dataset; a larger study would be required in the future to estimate wider clinical utility. A large degree of variation was observed across some classification statistics, potentially signifying a large degree of biological heterogeneity in the dataset. A potential cause for this could be the effect of inherent molecular heterogeneity of the tumour microenvironment; or perhaps varying extents of lymphocyte infiltration present in specimens. The difficulty of annotating samples is likely to introduce noise into the dataset; alongside inconsistencies in sample preparation and measurement procedure

## **1.5 Conclusion**

The use of FTIR in a clinical setting is still in its infancy, however the work covered here shows that it has the potential to be of significant benefit as a prognostic tool. The addition of ASMA information was shown to be beneficial in one case, and demonstrates that additional information from other modalities could lead to the creation of a novel and informative prognostic tool. FTIR spectroscopy has been shown to be capable of observing a molecular fingerprint associated with poor prognosis. It would be of considerable benefit to be able to direct patients with poor prognoses towards appropriate treatment. Currently it is not ethically possible to select patients for neo-adjuvant treatment in window trials; patients who would not benefit from such treatment would incur unnecessary adverse effects and additional health risks. It is currently only possible to determine patient prognoses after post-surgical nodal biopsies have taken place. The work presented here would allow for this crucial window of opportunity to be seized and to enable the development of new treatment methods



to take place.

## Bibliography

- [1] Matt Lechner, Jacklyn Liu, Liam Masterson, and Tim R. Fenton. HPV-associated oropharyngeal cancer: epidemiology, molecular biology and clinical management. *Nature Reviews Clinical Oncology*, 0123456789, 2022.
- [2] Head and neck cancers statistics, Sep 2021.
- [3] Mauricio E. Gamez, Ryan Kraus, Michael L. Hinni, Eric J. Moore, Daniel J. Ma, Stephen J. Ko, Jean Claude M. Rwigema, Lisa A. McGee, Michele Y. Halyard, Matthew R. Buras, Robert L. Foote, and Samir H. Patel. Treatment outcomes of squamous cell carcinoma of the oral cavity in young adults. *Oral Oncology*, 87(August):43–48, 2018.
- [4] Torbjörn Ramqvist and Tina Dalianis. An epidemic of oropharyngeal squamous cell carcinoma (OSCC) due to human papillomavirus (HPV) infection and aspects of treatment and prevention. *Anticancer research*, 31(5):1515–9, 2011.
- [5] Victor Lopes, Paul Murray, Hazel Williams, Ciaran Woodman, John Watkinson, and Max Robinson. Squamous cell carcinoma of the oral cavity rarely harbours oncogenic human papillomavirus. *Oral Oncology*, 47(8):698–701, 2011.
- [6] Athanassios Argiris, Neck Cancer Program, Michalis V Karamouzis, Neck Cancer Program, David Raben, Robert L Ferris, Neck Cancer Program, Cancer Immunology, and Immunoprevention Program. Head and neck cancer Athanassios. *Lancet*, 371(9625):1695–1709, 2008.

- [7] Conor P. Barry, Chetan Katre, Elena Papa, James S. Brown, Richard J. Shaw, Fazilet Bekiroglu, Derek Lowe, and Simon N. Rogers. De-escalation of surgery for early oral cancer-is it oncologically safe? *British Journal of Oral and Maxillofacial Surgery*, 51(1):30–36, 2013.
- [8] Ari J. Rosenberg and Everett E. Vokes. Optimizing Treatment De-Escalation in Head and Neck Cancer: Current and Future Perspectives. *The Oncologist*, 26(1):40–48, 2021.
- [9] Mahul B. Amin, Frederick L. Greene, Stephen B. Edge, Carolyn C. Compton, Jeffrey E. Gershenwald, Robert K. Brookland, Laura Meyer, Donna M. Gress, David R. Byrd, and David P. Winchester. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA: A Cancer Journal for Clinicians*, 67(2):93–99, 2017.
- [10] Sang Ik Park, Jeffrey P. Guenette, Chong Hyun Suh, Glenn J. Hanna, Sae Rom Chung, Jung Hwan Baek, Jeong Hyun Lee, and Young Jun Choi. The diagnostic performance of CT and MRI for detecting extranodal extension in patients with head and neck squamous cell carcinoma: a systematic review and diagnostic meta-analysis. *European Radiology*, 31(4):2048–2061, 2021.
- [11] Benjamin H. Kann, Daniel F. Hicks, Sam Payabvash, Amit Mahajan, Justin Du, Vishal Gupta, Henry S. Park, James B. Yu, Wendell G. Yarbrough, Barbara A. Burtneess, Zain A. Husain, and Sanjay Aneja. Multi-institutional validation of deep learning for pretreatment identification of extranodal extension in head and neck squamous cell carcinoma. *Journal of Clinical Oncology*, 38(12):1304–1311, 2020.
- [12] Rebekah K. O’Donnell, Michael Kupferman, S. Jack Wei, Sunil Singhal, Randal Weber, Bert O’Malley, Yi Cheng, Mary Putt, Michael Feldman,

- Barry Ziober, and Ruth J. Muschel. Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. *Oncogene*, 24(7):1244–1251, 2005.
- [13] Paul Roepman, Lodewyk F.A. Wessels, Nienke Kettelarij, Patrick Kemmeren, Antony J. Miles, Philip Lijnzaad, Marcel G.J. Tilanus, Ronald Koole, Gert Jan Hordijk, Peter C. Van Der Vliet, Marcel J.T. Reinders, Piet J. Slootweg, and Frank C.P. Holstege. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. *Nature Genetics*, 37(2):182–186, 2005.
- [14] D. S. Rickman, R. Millon, A. De Reynies, E. Thomas, C. Wasylyk, D. Muller, J. Abecassis, and B. Wasylyk. Prediction of future metastasis and molecular characterization of head and neck squamous-cell carcinoma based on transcriptome and genome analysis by microarrays. *Oncogene*, 27(51):6607–6622, 2008.
- [15] Chenguang Zhao, Huiru Zou, Jun Zhang, Jinhui Wang, and Hao Liu. An integrated methylation and gene expression microarray analysis reveals significant prognostic biomarkers in oral squamous cell carcinoma. *Oncology Reports*, 40(5):2637–2647, 2018.
- [16] Diana Graizel, Ayelet Zlotogorski-Hurvitz, Igor Tsesis, Eyal Rosen, Ron Kedem, and Marilena Vered. Oral cancer-associated fibroblasts predict poor survival: Systematic review and meta-analysis. *Oral Diseases*, 26(4):733–744, 2020.
- [17] Tseng-Cheng Chen, Chen-Tu Wu, Cheng-Ping Wang, Wan-Lun Hsu, Tsung-Lin Yang, Pei-Jen Lou, Jenq-Yuh Ko, and Yih-Leong Chang. Associations among pretreatment tumor necrosis and the expression of hif-1 $\alpha$  and pd-l1 in advanced oral squamous cell carcinoma and the prognostic impact thereof. *Oral Oncology*, 51(11):1004–1010, 2015.

- [18] Anthony C Nichols, Pencilla Lang, Eitan Prisman, Eric Berthelet, Eric Tran, Sarah Hamilton, Jonn Wu, Kevin Fung, John R de Almeida, Andrew Bayley, et al. Treatment de-escalation for hpv-associated oropharyngeal squamous cell carcinoma with radiotherapy vs. trans-oral surgery (orator2): study protocol for a randomized phase ii trial. *BMC cancer*, 20(1):1–13, 2020.
- [19] Alexandra Sala, David J. Anderson, Paul M. Brennan, Holly J. Butler, James M. Cameron, Michael D. Jenkinson, Christopher Rinaldi, Ashton G. Theakstone, and Matthew J. Baker. Biofluid diagnostics by FTIR spectroscopy: A platform technology for cancer detection. *Cancer Letters*, 477(December 2019):122–130, 2020.
- [20] Vera E. Sitnikova, Mariia A. Kotkova, Tatiana N. Nosenko, Tatiana N. Kotkova, Daria M. Martynova, and Mayya V. Uspenskaya. Breast cancer detection by ATR-FTIR spectroscopy of blood serum and multivariate data-analysis. *Talanta*, 214(October 2019):120857, 2020.
- [21] Daniela Lazaro-Pacheco, Abeer Shaaban, Gouri Baldwin, Nicholas Ak-inwale Titiloye, Shazza Rehman, and Ihtesham ur Rehman. Deciphering the structural and chemical composition of breast cancer using FTIR spectroscopy. *Applied Spectroscopy Reviews*, 0(0):1–15, 2020.
- [22] Qingbo Li, Can Hao, Xue Kang, Jialin Zhang, Xuejun Sun, Wenbo Wang, and Haishan Zeng. Colorectal Cancer and Colitis Diagnosis Using Fourier Transform Infrared Spectroscopy and an Improved K-Nearest-Neighbour Classifier. *Sensors*, 17(12):2739, 2017.
- [23] Sebastian Berisha, Mahsa Lotfollahi, Jahandar Jahanipour, Ilker Gurcan, Michael Walsh, Rohit Bhargava, Hien Van Nguyen, and David Mayerich. Deep learning for FTIR histology: leveraging spatial and spectral features with convolutional neural networks. *Analyst*, 144(5):1642–1653, 2019.

- [24] J. Ingham, M.J. Pilling, D.S. Martin, C.I. Smith, B.G. Ellis, C.A. Whitley, M.R.F. Siggel-King, P. Harrison, T. Craig, A. Varro, D.M. Pritchard, A. Varga, P. Gardner, P. Weightman, and S. Barrett. A novel FTIR analysis method for rapid high-confidence discrimination of esophageal cancer. *Infrared Physics and Technology*, 102, 2019.
- [25] Donna E. Maziak, Minh T. Do, Farid M. Shamji, Sudhir R. Sundaresan, D. Garth Perkins, and Patrick T.T. Wong. Fourier-transform infrared spectroscopic study of characteristic molecular structure in cancer cells of esophagus: An exploratory study, 2007.
- [26] A. L. M. Batista de Carvalho, M. Pilling, P. Gardner, J. Doherty, G. Cinque, K. Wehbe, C. Kelley, L. A. E. Batista de Carvalho, and M. P. M. Marques. Chemotherapeutic response to cisplatin-like drugs in human breast cancer cells probed by vibrational microspectroscopy. *Faraday Discuss.*, 187:273–298, 2016.
- [27] J Dhanda, A Triantafyllou, T Liloglou, H Kalirai, B Lloyd, R Hanlon, R J Shaw, and D R Sibson. SERPINE1 and SMA expression at the invasive front predict extracapsular spread and survival in oral squamous cell carcinoma. (August):2114–2121, 2014.
- [28] Marilena Vered, Alex Dobriyan, Dan Dayan, Ran Yahalom, Yoav P. Talmi, Lev Bedrin, Iris Barshack, and Shlomo Taicher. Tumor-host histopathologic variables, stromal myofibroblasts and risk score, are significantly associated with recurrent disease in tongue cancer. *Cancer Science*, 101(1):274–280, 2010.
- [29] M. G. Kellermann, L. M. Sobral, S. D. Da Silva, K. G. Zecchin, E. Graner, M. A. Lopes, I. Nishimoto, L. P. Kowalski, and R. D. Coletta. Myofibroblasts in the stroma of oral squamous cell carcinoma are associated with poor prognosis [3]. *Histopathology*, 51(6):849–853, 2007.

- [30] The GIMP Development Team. Gimp.
- [31] A. Köhler, J. Sulé-Suso, G. D. Sockalingum, M. Tobin, F. Bahrami, Y. Yang, J. Pijanka, P. Dumas, M. Cotte, D. G. Van Pittius, G. Parkes, and H. Martens. Estimating and correcting Mie scattering in synchrotron-based microscopic fourier transform infrared spectra by extended multiplicative signal correction. *Applied Spectroscopy*, 62(3):259–266, 2008.
- [32] Pitard F.F. An Introduction to the Theory of Sampling: An Essential Part of Total Quality Management. *Comprehensive Chemometrics*, pages 1–16, 2009.
- [33] Wolfgang Karl Härdle and Léopold Simar. *Applied multivariate statistical analysis, fourth edition*. 2015.
- [34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [35] Cameron Davidson-Pilon. lifelines: survival analysis in python. *Journal of Open Source Software*, 4(40):1317, 2019.
- [36] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and

- SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.
- [37] Félix-Antoine Fortin, François-Michel De Rainville, Marc-André Gardner, Marc Parizeau, and Christian Gagné. DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, 13:2171–2175, jul 2012.
- [38] Simon N. Rogers, James S. Brown, Julia A. Woolgar, Derek Lowe, Patrick Magennis, Richard J. Shaw, David Sutton, Douglas Errington, and David Vaughan. Survival following primary surgery for oral cancer. *Oral Oncology*, 45(3):201–211, 2009.
- [39] S Sawyer. The Greenwood and Exponential Greenwood Confidence Intervals in Survival Analysis. *Health (San Francisco)*, (i):2–5, 2003.
- [40] Alhadi Almangush, Ilkka Heikkinen, Antti A. Mäkitie, Ricardo D. Coletta, Esa Läärä, Ilmo Leivo, and Tuula Salo. Prognostic biomarkers for oral tongue squamous cell carcinoma: A systematic review and meta-analysis. *British Journal of Cancer*, 117(6):856–866, 2017.
- [41] Richard J. Shaw, Derek Lowe, Julia A. Woolgar, James S. Brown, E. David Vaughan, Christopher Evans, Huw Lewis-Jones, Rebecca Hanlon, Gillian L. Hall, and Simon N. Rogers. Extracapsular spread in oral squamous cell carcinoma. *Head Neck*, 36(10):NA–NA, 2009.
- [42] Maxime Mermoud, Genrich Tolstonog, Christian Simon, and Yan Monnier. Extracapsular spread in head and neck squamous cell carcinoma: A systematic review and meta-analysis. *Oral Oncology*, 62:60–71, 2016.
- [43] David Chin, Glen M. Boyle, Rebecca M. Williams, Kaltin Ferguson, Nirmala Pandeya, Julie Pedley, Catherine M. Campbell, David R. Theile, Peter G. Parsons, and William B. Coman. Novel markers for poor prognosis

- in head and neck cancer. *International Journal of Cancer*, 113(5):789–797, 2005.
- [44] Chun Ta Liao, Li Yu Lee, Chuen Hsueh, Chien Yu Lin, Kang Hsing Fan, Hung Ming Wang, Chia Hsun Hsieh, Shu Hang Ng, Chih Hung Lin, Chung Kan Tsao, Chung Jan Kang, Tuan Jen Fang, Shiang Fu Huang, Kai Ping Chang, Lan Yan Yang, and Tzu Chen Yen. Pathological risk factors stratification in pN3b oral cavity squamous cell carcinoma: Focus on the number of positive nodes and extranodal extension. *Oral Oncology*, 86(September):188–194, 2018.
- [45] Christine T. Lwin, Rebecca Hanlon, Derek Lowe, James S. Brown, Julia A. Woolgar, Asterios Triantafyllou, Simon N. Rogers, Fazilet Bekiroglu, Huw Lewis-Jones, Hulya Wiesmann, and Richard J. Shaw. Accuracy of MRI in prediction of tumour thickness and nodal stage in oral squamous cell carcinoma. *Oral Oncology*, 48(2):149–154, 2012.
- [46] Peng Wang, Aron Popovtzer, Avraham Eisbruch, and Yue Cao. An approach to identify, from DCE MRI, significant subvolumes of tumors related to outcomes in advanced head-and-neck cancer. *Medical Physics*, 39(8):5277–5285, 2012.
- [47] Tobias Waech, Shila Pazahr, Vittoria Guarda, Niels J. Rupp, Martina A. Broglie, and Grégoire B. Morand. Measurement variations of MRI and CT in the assessment of tumor depth of invasion in oral cancer: A retrospective study. *European Journal of Radiology*, 135, 2021.
- [48] Jenny K. Hoang, Jyotsna Vanka, Benjamin J. Ludwig, and Christine M. Glastonbury. Evaluation of cervical lymph nodes in head and neck cancer with CT and MRI: Tips, traps, and a systematic approach. *American Journal of Roentgenology*, 200(1):17–25, 2013.
- [49] Ming Hui Mao, Shu Wang, Zhi en Feng, Jin Zhong Li, Hua Li, Li zheng



- Qin, and Zheng xue Han. Accuracy of magnetic resonance imaging in evaluating the depth of invasion of tongue cancer. A prospective cohort study. *Oral Oncology*, 91(October 2018):79–84, 2019.
- [50] Izabela Zawlik, Ewa Kaznowska, Jozef Cebulski, Magdalena Kolodziej, Joanna Depciuch, Jitraporn Vongsvivut, and Marian Cholewa. FPA-FTIR Microspectroscopy for Monitoring Chemotherapy Efficacy in Triple-Negative Breast Cancer. *Scientific Reports*, 6:1–8, 2016.
- [51] Holly J. Butler, Paul M. Brennan, James M. Cameron, Duncan Finlayson, Mark G. Hegarty, Michael D. Jenkinson, David S. Palmer, Benjamin R. Smith, and Matthew J. Baker. Development of high-throughput ATR-FTIR technology for rapid triage of brain cancer. *Nature Communications*, 10(1):1–9, 2019.
- [52] J Zapala, J Czopek, A Lazar, and R Tomaszewska. Proliferative index activity in oral squamous cell carcinoma : indication for postoperative radiotherapy ? pages 1189–1194, 2014.
- [53] Thomas Scholzen and Johannes Gerdes. The Ki-67 Protein : From the Known and. 322(August 1999):311–322, 2000.
- [54] Neelam G Shah, Trupti I Trivedi, Rajen A Tankshali, Jignesh V Goswami, Dhaval H Jetly, Shilin N Shukla, Pankaj M Shah, and Ramtej J Verma. Prognostic significance of molecular markers in oral squamous cell carcinoma : a multivariate analysis. (December):1544–1556, 2009.
- [55] César Rivera, Ana Karina Oliveira, Rute Alves Pereira Costa, Tatiane De Rossi, and Adriana Franco Paes Leme. Prognostic biomarkers in oral squamous cell carcinoma: A systematic review. *Oral Oncology*, 72:38–47, 2017.