The Application of Artificial Intelligence and Image Analysis to Novel Prognostic Classification Systems of Colorectal Cancer

By

Christopher J. Bigley

BSc (Hons)

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

To

The School of Cancer Sciences

College of Medical, Veterinary and Life Sciences

The University of Glasgow

September 2022

The work presented in this thesis was performed by the author except where acknowledged. This thesis has not been submitted for a degree or diploma at this or any other institution.

Christopher James Bigley

September 2022

Acknowledgements

Firstly, I would like to thank my primary supervisor, Dr Antonia Roseweir, and my second supervisor, Professor Joanne Edwards. Your expertise, mentorship, guidance, and unwavering support has carried me through this process. Working with the both of you over the past four years has been an immense pleasure and I cannot begin to thank you enough for seeing me through my PhD.

Thank you to my family, my mother Alison, this thesis could never have been realised without your unparalleled expertise in digital pathology as well as constant emotional support throughout everything, my father Graham and my brother Michael, thank you for always believing in me, I hope I have made you proud. And to my now extended family, Leah and Teddy, thank you for always supporting me and welcoming me so warmly.

To everyone in the Edwards Lab, thank you for your support, comic relief and keeping me grounded, you are the reason I came in every day, and I do not believe I will ever work with such an enjoyable group of people again.

Thank you to Dr John Waller and everyone at OracleBio for your support on all things technical and making me feel welcome when I was there.

Thank you to Professor Karin Oien not only for your pathology tutorship, but your pastoral support when it was needed most.

To all my friends both in Glasgow and afar, thank you for all the pub – based group therapy sessions.

Finally to Hannah, it would take more than this thesis to detail everything you have done, and continue to do for my life, you are my rock, everything I do is for you and for us. Thank you.

Summary

Colorectal cancer (CRC) is the third most common form of cancer in the world and the second most common cancer related mortality. Adequate staging of CRC is important for understanding patient prognosis and determining appropriate therapy regimens. CRC staging is currently performed according to the Tumour, Node, Metastasis (TNM) staging criteria, which has remained the gold standard around the world since its introduction. However, the variable prognosis of Stage II / node negative disease and uncertainty around best therapeutic practices for these patients has been a continuing issue with the TNM system, one that is still yet to be adequately addressed. Through extensive research, novel prognostic features, assessed on diagnostic Haematoxylin & Eosin (H&E) sections and through simple Immunohistochemistry (IHC), have been shown to supplement the TNM criteria and address this unmet clinical need. Furthermore, novel classification systems that incorporate multiple features of the Tumour Microenvironment (TME) and assign patients to independent groups have been developed and often stratify patients for prognosis better than the TNM system, as well as providing additional prognostic and theragnostic information. The adoption of these novel factors and classification systems into clinical pathology has been hindered by persistent interobserver variability and a lack of clear and standardised assessment criteria.

Image analysis presents a means to reduce the subjectivity of these criteria and increase their potential clinical utility. With the advent of artificial intelligence and its continued development within image analysis, the automated assessment of novel features on clinical sections is becoming increasingly reliable and reproducible. Therefore, this thesis aims to utilise image analysis and artificial intelligence to automate the assessment of specific features of the CRC TME, assess the prognostic utility of novel TME features individually and in combination, and compare the performance of digital assessment to human assessment.

The Glasgow Microenvironment Score (GMS) is a combined assessment of the stromal density of the tumour, quantified through the Tumour Stroma Percentage (TSP), and the peritumoural inflammatory response, assessed using the Klintrup – Mäkinen (KM) criteria, that assigns patients to one of three individually prognostic groups. Using a Convolutional Neural Network (CNN) to semantically segment H&E Whole Slide Images (WSI) and quantify the tumour associated stroma demonstrated that image analysis is able to reliably conduct TSP assessment across multiple retrospective patient cohorts and a large clinical trial cohort, prognostically stratify these patients according to the TSP criteria, and outperform human assessment for prognostic significance. Image analysis quantification of

peritumoural lymphocyte density on H&E WSI with manually annotated invasive margins showed a significant association with prognosis, comparable to that seen in manual KM assessment, again across multiple cohorts. Combining both image analysis approaches according to the GMS criteria outperformed pathologist assessment for survival stratification, highlighting the ability of image analysis algorithms to reliably perform individual assessments and retain the prognostic significance when used in combination.

The Phenotypic Subtypes of CRC are a translation of the phenotypic signatures of the Consensus Molecular Subtypes (CMS) to tissue – based assessment, incorporating Ki67 IHC into the GMS criteria with the TSP and KM. Ki67 expression is utilised to further stratify the GMS group with intermediate prognosis, providing additional information about the TME. To quantify Ki67 expression, a CNN was again used to semantically segment Tissue MicroArray (TMA) cores stained for Ki67 via IHC and the percentage of Ki67+ tumour cells was determined using an automated, CNN – based cell detection algorithm. Ki67 expression determined through automated analysis was significantly associated with prognosis individually, and when combined with the TSP and KM criteria, the Phenotypic Subtypes determined through image analysis were highly prognostic again across multiple cohorts. Furthermore, the image analysis subtypes identified a group of patients with a chemotherapy dependent improvement in survival, demonstrating the clinical utility of image analysis for determining patient prognosis and potentially guiding therapy regimens.

The data presented in the current thesis demonstrates that image analysis is able to reliably and reproducibly assess novel features of the TME from clinical WSI, perform these assessments across multiple independent patient cohorts, significantly stratify patients for prognosis, and has the potential to be utilised in clinical pathology to aid therapeutic decisions and improve patient outcomes.

List of Figures

List of Tables

Table 6.4 Relationship between Phenotypic Subtype, Clinicopathological Features and DFS in TransSCOT Cohort

Abbreviations

AI Artificial Intelligence

AJCC American Joint Committee on Cancer

APC Adenomatous Polyposis Coli

APP Analysis Protocol Package

AUC Area Under Curve

CAD Computer Aided Diagnosis

CAF Cancer Associated Fibroblast

CAP College of American Pathologists

CAPOX Capecitabine

CD Cluster of Differentiation

CI Confidence Interval

CIMP CpG Island Methylator Phenotype

CIN Chromosomal Instability

CMS Consensus Molecular Subtypes

CNN Convolution Neural Network

CP Carcinoma Percentage

CpG Cytosine Phosphodiester Bond Guanine

CRC Colorectal Cancer

CRCSC Colorectal Cancer Subtyping Consortium

CRIS Colorectal Cancer Intrinsic Subtypes

CSS Cancer Specific Survival

DAB 3, 3’ – Diaminobenzidine

DFS Disease Free Survival

DL Deep Learning

DP Digital Pathology

ECM Extracellular Matrix

EGFR Epidermal Growth Factor Receptor

EMT Epithelial Mesenchymal Transition

ER Oestrogen Receptor

ERK Extracellular Signal – Regulated Kinase

FAP Familial Adenomatous Polyposis

FDA United States Food and Drug Administration

FOLFOX Folinic Acid, 5 – Fluorouracil, Oxaliplatin

FOV Field of View

GCHP Goblet Cell Hyperplastic Polyp

GDP Guanosine Diphosphate

gFOBT guaiac Faecal Occult Blood Test

GMS Glasgow Microenvironment Score

GPU Graphics Processing Unit

Grad – CAM Gradient – Weighted Class Activation Mapping

GTP Guanosine Triphosphate

H&E Haematoxylin and Eosin

HER2 Human Epidermal Growth Factor Receptor 2

HDI Human Development Index

HNPCC Hereditary Non – Polyposis Colorectal Cancer

HP Hyperplastic Polyp

HR Hazard Ratio

IF Immunofluorescence

IGF Insulin Like Growth Factor

IHC Immunohistochemistry

KM Klintrup – Mäkinen

KRAS Kirsten Rat Sarcoma Virus

LoG Laplacian of Gaussian

LS Lynch Syndrome

MAPK Mitogen Activated Protein Kinase

ML Machine Learning

MMR Mismatch Repair

MPHP Mucin Poor Hyperplastic Polyp

MSI Microsatellite Instability

MSS Microsatellite Stable

MVHP Microvesicular Hyperplastic Polyp

NHS National Health Service

NK Natural Killer Cells

NPIC National Pathology Imaging Co – operative

OS Overall Survival

PDX Patient Derived Xenograft

PI Positivity / Proliferative Index

PR Progesterone Receptor

qFIT quantitative Faecal Immunochemical Test

RCPath Royal College of Pathologists

ReLU Rectified Liner Unit

RFS Recurrence Free Survival

ROC Receiver Operator Characteristic

ROI Region of Interest

SCNA Somatic Copy Number Alteration

SRCC Signet Ring Cell Carcinoma

SSA Sessile Serrated Adenoma

SVM Support Vector Machine

TGFß Transforming Growth Factor ß

TIL Tumour Infiltrating Lymphocyte

Tis Tumour *in situ*

TMA Tissue Microarray

TME Tumour Microenvironment

TNM Tumour, Node, Metastasis

TSA Traditional Serrated Adenoma

TSG Tumour Suppressor Gene

TSP Tumour Stroma Percentage

TSR Tumour Stroma Ratio

UICC Union for International Cancer Control

WSI Whole Slide Image

1. Introduction

1.1 Colorectal Cancer Epidemiology

Colorectal Cancer (CRC) is the third most common form of cancer in the world, with 1,148,515 new cases arising in 2020, accounting for 10.6% of cancer incidence in men; the third most common after lung and prostate, and 9.4% of cancer incidence in women; the second most common after breast (Sung et al., 2021). In both sexes worldwide, CRC accounts for 9.4% of all cancer related mortalities, totalling ~930,600 deaths in 2020 (Sung et al., 2021). CRC incidence and mortality vary between socioeconomic regions across the world, with a notably higher incidence in developed countries. In countries with a high or very high Human Development Index (HDI), incidence of CRC is ~24.5 cases per 100,000 of the population, however, this is markedly lower in countries with a medium or low HDI, with an incidence of ~6.4 cases per 100,000 of the population. Some regions with a historically low and medium HDI, Eastern Europe, Asia, and South America most notably, have shown an increasing CRC incidence since 1980, likely due to the adoption of western lifestyle factors such as lack of exercise leading to obesity, and tobacco and alcohol consumption (Vabi et al., 2021).

In the UK, CRC is the third most common cancer and second most common cause of cancer death representing around 11% of all cancer incidence with 42,885 diagnoses between 2016 – 2018, 56% of which occurred in males, and 16,807 cancer deaths, 55% of which occurred in males (Caul & Broggio, 2017). In Scotland, a bowel screening program was piloted between 2000 – 2007 and fully rolled out to the whole population aged 50 – 74 in 2009, in which guaiac faecal occult blood tests (gFOBT) were performed biennially on the eligible population. The gFOBT test was however later replaced with a quantitative faecal immunochemical test (qFIT) as it demonstrated greater detection sensitivity, a patient with a positive result from either test was then referred for colonoscopy to identify polyps or CRC. Between 2000 – 2007, 1,487,999 individuals participated at least once in the screening programme with an average uptake of 55%, and 24,817 CRCs were diagnosed through screening and non – screening pathways. In the screening age range of 50 – 74 years, incidence decreased from 154.4 cases per 100,000 in 2000, the first year the bowel screening programme was piloted in three of the fourteen Scottish NHS Health Boards, to 123.9 cases per 100,000 in 2017, however, an increase in incidence in individuals aged <50 years was observed in the same time period, from 5.3 cases per 100,000 in 2000 to 6.8 cases per 100,000 in 2017 (Clark et al., 2020).

1.2 Colorectal Cancer Pathogenesis

The mechanisms underpinning sporadic CRC carcinogenesis are complex and heterogeneous, involving multiple genetic pathways and cellular processes. The key driver of the development of CRC is genomic instability, which arises through three distinct pathways: the Chromosomal Instability (CIN) pathway, the Microsatellite Instability (MSI) pathway, and the CpG Island Methylator Phenotype (CIMP).

1.2.1 Chromosomal Instability Pathway

Fearon and Vogelstein first proposed the multi – hit hypothesis of CRC carcinogenesis in 1990, detailing the accumulation of genetic and epigenetic alterations that contribute to the initiation, promotion, and progression of the adenoma – carcinoma sequence, which gives rise to 65 – 75% of CRC carcinomas (Fearon & Vogelstein, 1990). The adenoma – carcinoma sequence describes the relatively linear sequence of mutational events occurring in the colonic epithelium, that initially leads to the formation of pre – malignant precursor lesions, most commonly adenomatous polyps, and then over the course of more than 10 years leads to the development of CRC (Jass, 2007). An early event in this sequence is loss of the Adenomatous Polyposis Coli (APC) tumour suppressor gene through inactivating somatic mutation, which occurs in 70 – 80% of sporadic CRC (Kinzler & Vogelstein, 1996), or loss of heterozygosity of chromosome 5q, which is reported in 30 – 40% of CRC tumours (Rowan et al., 2000). This loss of APC function results in constitutive activation of the canonical Wnt signalling pathway, of which APC is a key negative regulator, leading to up – regulation of downstream oncogenic targets such as Myc and Cyclin D1 which have roles in proliferation and apoptosis (He et al., 1998; L. Zhang & Shay, 2017). The next key oncogenic event in the CIN sequence, generally thought to occur during or immediately preceding late adenoma development, is activating mutations of the K – *ras* oncogene, which are present in 35 – 42% of sporadic CRCs (Vogelstein et al., 1988). KRAS is a GTP binding protein involved primarily in EGFR signal transduction through the Raf – Mitogen Activated Protein Kinase (MAPK) pathway, resulting in cellular proliferation and survival (Fernández-Medarde & Santos, 2011). Activating mutations of KRAS inhibit the innate GTPase activity that regulates the conversion of active KRAS – GTP to inactive KRAS – GDP, leading to constitutively sustained MAPK signalling (Liu et al., 2021; Uprety & Adjei, 2020). The last major event in the adenoma – carcinoma sequence, is mutation of the *TP53* gene, the most frequently altered gene in human cancers. The p53 protein is involved in control of cell cycle arrest, facilitating DNA damage repair, and inducing

apoptosis when necessary, therefore its functional inactivation has clear implications for carcinogenesis (Armaghany et al., 2012). The finding that p53 alteration or 17p allelic loss is present in only 4 – 26% of adenomas but 50 – 75% of carcinomas has established the belief that this is the initiating for factor of the transition from precursor lesion to invasive carcinoma and that functional inactivation of p53 occurs around this time (Leslie et al., 2002).

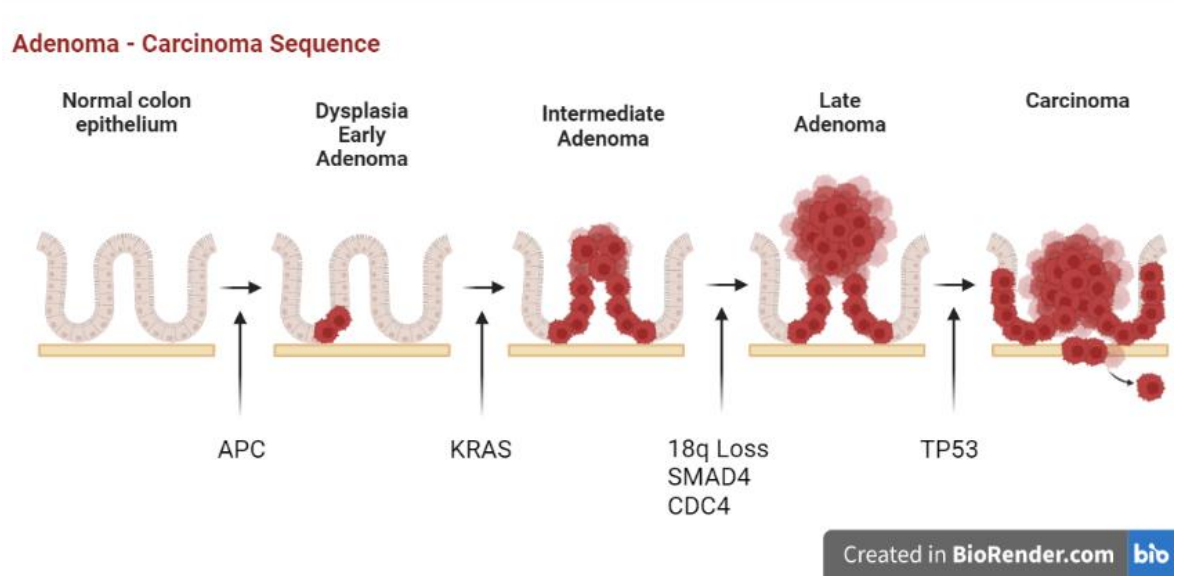


Figure 1.1 The Adenoma – Carcinoma Sequence. *The chromosomal instability pathway or adenoma – carcinoma sequence of CRC development proposed by Fearon & Vogelstein, detailing disease progression from normal colonic mucosa to carcinoma and the genetic alteration landmarks promoting this transition.*

1.2.2 Microsatellite Instability Pathway

During normal DNA replication, errors in nucleotide assembly by DNA polymerase are detected and corrected by the Mismatch Repair (MMR) machinery, of which there are at least seven constituent proteins in human cells (Bateman, 2021). In CRC, genetic alterations in four key MMR machinery protein coding genes; MLH1, MSH2, MSH6, and PMS2, results in a breakdown of the normal MMR processes, causing genomic instability that contributes to disease progression. This can occur through somatic mutation or in the case of MLH1, hypermethylation of the gene promoter (Kane et al., 1997). Clinical detection of MMR deficient (dMMR) tumours is conducted via immunohistochemistry (IHC) for the four MMR proteins, with dMMR tumours showing complete loss of expression of at least one. Deficiency in the MMR process leads to the expansion or contraction of Short Tandem Repeats (STRs) known as microsatellites; stretches of hundreds of nucleotides of mono, di, tri, and tetranucleotide repeats, that is the result of strand slippage and DNA polymerase

stutter, which is passed on to the daughter cells following mitosis (Ionov et al., 1993). This change in microsatellite repeat length is termed Microsatellite Instability (MSI). Diagnosis of MSI is performed through Polymerase Chain Reaction (PCR) testing of a panel of 5 gene loci, known as the Bethesda Panel: two mononucleotide loci – BAT – 25 & BAT – 26, and three dinucleotide loci – D2S123, D5S346, & D17S250, with tumours harbouring alterations at ≥2 of these loci classified as MSI – High (Murphy et al., 2006). Though MSI initially contributes to the progression of the disease, stage I – III patients with MSI tumours have a significantly better disease specific 5 – year prognosis than those with microsatellite stable (MSS) tumours (Benatti et al., 2005; Popat et al., 2005a).

1.2.3 CpG Island Methylator Phenotype Pathway

Epigenetic regulation of gene expression involves modifications to the DNA strand without alteration to the nucleotide sequence. Silencing of tumour suppressor gene function via aberrant hypermethylation of gene promoter regions is common in CRC, accounting for ~20% of sporadic tumours, and involves the binding of methyl groups to cytosine residues in areas rich in cytosine repeats. These areas are known as CpG islands, so called due to the presence of long dinucleotide repeats of Cytosine (C) and Guanine (G) residues linked by a phosphodiester bond (p), which are normally maintained in a non – methylated state permitting gene expression (Bird, 1986; Issa, 2004). Patients harbouring tumours with prominent CpG island hypermethylation are termed as possessing the CpG Island Methylator Phenotype (CIMP) – positive and are diagnosed as such through positive methylation detection at a panel of specific gene loci, originally MINT1, MINT2, MINT31, CDKN2A, and hMLH1 though subsequent studies have suggested expanding upon this panel to include more genetic loci, granting a more complete overview of the tumour's methylation status (Chan et al., 2002; Weisenberger et al., 2006). Dependent upon the panel used and the methylation status findings, CRC tumours can be categorized into three CIMP epigenotype groups, CIMP – High, CIMP – Low, and CIMP – Negative (Kaneda & Yagi, 2011).

It has long been understood that CIMP – High or CIMP – 1 tumours are characterized by transcriptional inactivation of the MMR gene *MLH1* through promoter hypermethylation and show an extremely strong association with activating BRAF mutations, predominantly $BRAF^{V600E}$ (Toyota et al., 1999; Weisenberger et al., 2006). This is due to sustained MEK / ERK signalling mediated by constitutively active $BRAF^{V600E}$ resulting in phosphorylation of MAFG, a transcriptional repressor that when bound to DNA recruits a co – repressor

complex leading to promoter methylation. ERK directed phosphorylation of MAFG prevents polyubiquitination and subsequent proteasomal degradation, facilitating MAFG mediated hypermethylation of the *MLH1* promoter region (Fang et al., 2014). Given the myriad of clinical factors associated with CIMP that can influence patient survival, such as an increased immunological response, the effect of the genes silenced by promoter hypermethylation, and the inconsistency in criteria with which CIMP is defined, no definitive association of CIMP with prognosis has yet been elucidated. However, meta – analyses have identified a general trend towards CIMP – High tumours conferring a poorer prognosis than CIMP – Low tumours (Advani et al., 2018; Jia et al., 2016).

1.2.4 Serrated Pathway

Neoplastic serrated polyps are morphologically distinct entities that were long believed to be benign, but it is now understood that these lesions harbour malignant potential. Carcinomas arising from these lesions, through what is termed the serrated pathway, may account for up to 20% of sporadic CRC cases and can possess a morphology similar to that of their precursor lesion, known as serrated adenocarcinoma. However, most tumours arising from serrated polyps do not resemble their precursor lesion and are characterised through their molecular signatures (Jass & Smith, 1992; Mäkinen, 2007; Yamane et al., 2014). These precursor lesions, known as serrated adenomas, are morphologically characterised by a saw – toothed appearance of the epithelial glands and can be histologically classified into three categories: hyperplastic polyps (HP), sessile serrated adenomas (SSA) and traditional serrated adenomas (TSA) (Pai et al., 2019). Furthermore, three morphologically distinct HPs are recognized: Microvesicular HP (MVHP), Goblet Cell HP (GCHP), and Mucin Poor HP (MPHP) (Torlakovic et al., 2003). The molecular pathogenesis of these lesions is notably different from those of the traditional adenoma – carcinoma sequence, which routinely harbour and may indeed be initiated by APC mutations, whereas APC mutations are relatively rare in the serrated pathway. Activating mutations in the *BRAF* gene, namely V600E, are present in ~80% of SSAs and are frequently associated with high levels of CIMP, the latter of which is considered the driving factor towards malignancy in the serrated pathway (O'Brien et al., 2006). While *BRAF* mutations are also present in 20 – 40% of TSAs, the primary driving mutations are in the *KRAS* gene, found in 50 – 70% of TSAs, which are also found in concurrence with CIMP though at lower levels than those found in SSAs (Wiland et al., 2014). In MVHP, the predominant driver mutation is again $BRAF^{V600E}$ occurring in ~70% of these polyps, with *KRAS* mutations occurring less frequently, ~15%

(S. Yang et al., 2004). An inverse incidence is seen in GCHP, where ~20% harbour *BRAF* mutations but ~45% harbour *KRAS* mutations. Due to their frequent occurrence with activating mutations in prominent oncogenes, carcinomas arising from the serrated pathway typically confer a significantly worse 5 – year prognosis than conventional carcinomas (García-Solano et al., 2010).

1.2.5 Hereditary Colorectal Cancer

Sporadic CRC, accounting for 70 – 80% of cases, is influenced by the interaction of the colonic environment and the genetic constitution of the patient with modifiable risk factors such as diet, smoking, alcohol consumption, and sedentary lifestyle leading to obesity. The remaining 20 – 30% of cases arise from genetically inherited disorders, of which two are particularly noteworthy, Familial Adenomatous Polyposis and Lynch Syndrome.

1.2.5.1 Familial Adenomatous Polyposis

Familial Adenomatous Polyposis (FAP) is an inherited, autosomal dominant, colorectal cancer predisposition syndrome accounting for less than 1% of CRC cases and is caused by deleterious germline mutations, of which ~800 have been identified (Béroud & Soussi, 1996), in the *APC* gene located on chromosome 5q21 – 22. There are two characterized forms of FAP both with identical genetic causes: classic and attenuated. Classic FAP is characterized by the development of hundreds to thousands of colorectal adenomata at diagnosis and beginning in early adolescence, with a near certain incidence of CRC by the late fourth decade of life (Bussey, 1975). Attenuated FAP is a less severe form of the disorder, with patients developing an average of 30 adenomata at diagnosis, though still with a highly variable range that can go into the high hundreds, and conferring an average CRC incidence risk of 69% by age 80 (Burt et al., 2004). It is additionally hypothesized that around 25% of FAP cases are the result of *de novo* germline mutations in the APC gene (Zeichner et al., 2012). Treatment of FAP initially revolves around prophylactic surgical intervention which, depending on specific phenotype and presentation, is performed between the ages of 15 – 25 for classical FAP patients but can be later for attenuated FAP patients. There are three surgical procedures undertaken for FAP cases: total proctocolectomy, colectomy with an ileorectal anastomosis, and restorative proctocolectomy (Guilherme Campos, 2014). The chosen procedure is dependent upon the patient's genetic lineage and colonoscopic findings, but all involve the removal of as much colonic mucosa

as possible coupled with thorough follow up to prevent the risk of additional adenomata and subsequent CRC. In addition to FAP, another inherited polyposis involving germline mutations of the *MUTYH* gene termed MAP has been identified, with patients averaging diagnosis at 48 years of age and presenting between 10 and 500 polyps at diagnosis. Although 60% of MAP patients are diagnosed with CRC at first presentation, the penetrance is 43% at 60 years of age (M. Nielsen et al., 2011).

1.2.5.2 Lynch Syndrome

The second key hereditary predisposition syndrome conferring susceptibility to CRC is Lynch Syndrome (LS), formerly known as Hereditary Non – Polyposis Colorectal Cancer (HPNCC). LS is the autosomal dominant inheritance of pathogenic germline mutations in one of the four genes encoding the DNA MMR machinery: *MLH1, MSH2, MSH6, and PMS2,* and accounts for 2 – 5% of CRC cases with a lifetime risk of developing CRC of 70 – 80% generally occurring before the age of fifty (Stoffel et al., 2010). While the majority of LS associated cancer diagnoses are of CRC, LS confers susceptibility to a variety of malignancies, including those occurring in additional sites of the GI tract, neurological cancers such as astrocytoma and oligodendroglioma, melanoma, and sex specific cancers such as endometrial, ovarian, and prostate (Helder-Woolderink et al., 2016; Koornstra et al., 2009; Raymond et al., 2013; Ryan et al., 2019; Therkildsen et al., 2015). Given that LS is a cancer predisposition syndrome and not a hereditary polyposis, wherein diagnosis can be made at the occurrence of pre – malignant lesions and prophylactic interventions with follow up colonoscopies can be implemented, diagnosis of LS is generally made synchronously to the diagnosis of CRC. Loss of immunohistochemical staining for one of the MMR proteins indicates LS, however, it is estimated that between 3 – 12% of LS associated CRC present with MMR competent tumours through IHC (Bartley et al., 2012). Diagnosis is further indicated through additional mutational testing, such as for $BRAF^{V600E}$ which is generally only present in sporadic CRC (Capper et al., 2013), and confirmed through sequencing of specific gene loci to identify the pathological variant (Berg et al., 2009) In the setting of CRC, colonic tumours and rectal tumours are diagnostically distinct for LS as sporadic MMR deficient colonic tumours are predominantly right sided, sporadic MMR deficient rectal lesions are by comparison extremely rare. Therefore, they are recommended to be assumed to be LS and undergo immediate germline testing with subsequent lifelong follow up for colonic and extracolonic malignancies if LS diagnosis is positive (De Rosa et al., 2016).

1.3 Colorectal Pathology

1.3.1 Gross Anatomy of the Large Bowel

The large bowel is the terminal organ of the digestive system and gastrointestinal tract, it begins at the terminus of the small intestine and ends at the rectum, encompassing an average of 5’ in length. The main function of the large bowel is water absorption and peristaltic movement of faeces towards expulsion at the anus. The origin of the large bowel that joins to the small intestine is the caecum, attached at the ileocaecal valve, around 1” inferior to which is the appendix. Superior to the caecum is the ascending (right) colon which rises to the hepatic (right colic) flexure where it turns 90° to become the transverse colon. The transverse colon is the longest segment of the colon, averaging ~20” in length and forming an arc from the hepatic flexure to the splenic (left colic) flexure, where it again turns 90° and proceeds inferiorly as the descending (left) colon. The terminal segment of the descending colon is the sigmoid colon, an “S” shaped section of the bowel that connects to the rectum, where the organ terminates at the anus (Figure 1.2).

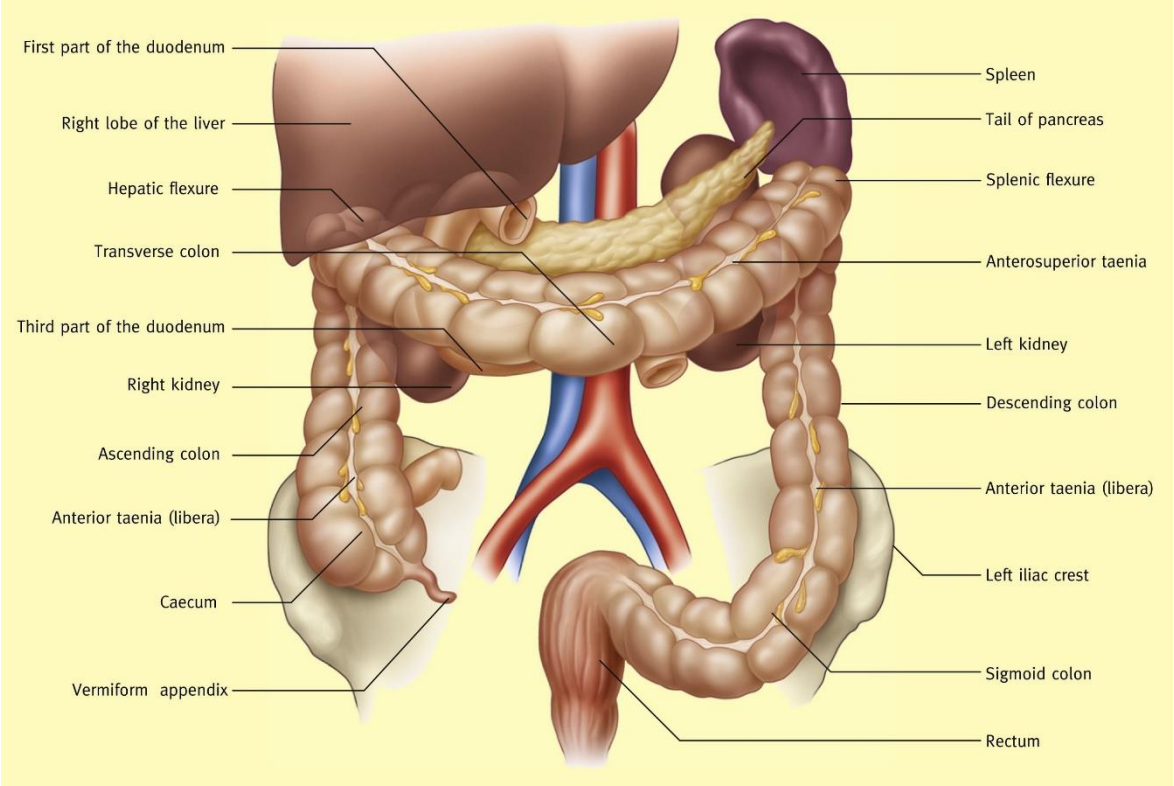


Figure 1.2 Anatomy of the Large Bowel. *Gross anatomy of the colon and rectum detailing the subdivisions, important anatomical structures, and global anatomical orientation. Adapted from (Ellis & Mahadevan, 2014).*

1.3.2 Site of Primary Tumour Development

The gross anatomy of the large bowel plays a key role in the pathogenesis and progression of CRC. Beginning in 1990, discordance in incidence and prognosis between right sided and left sided disease led to the routine reporting of anatomical site of disease at diagnosis of CRC (Bufill, 1990). Since this study, research into the clinical importance of primary tumour site of CRC has elucidated the distinct differences in pathogenesis, prognosis, and management of right vs left sided tumours to the extent they may now be considered distinct pathological entities. This distinction may originate during development as the anatomical sites have different embryological origins; the proximal (right) colon develops from the midgut and the distal (left) colon and rectum from the hindgut (Sadler, 2018). Epidemiologically, there is a greater incidence of left sided CRC than right sided CRC, with a 30 – year study in the US demonstrating nearly a 10% greater occurrence of left sided disease, however, there are conflicting arguments as to the prognostic difference between the disease sites which could be dependent upon the stage at diagnosis (Cheng et al., 2011; Patel et al., 2018; C. B. Wang et al., 2019; Warschkow et al., 2016). Site dependent pathological assessment of CRC highlights that right sided disease is generally associated with factors indicating poorer prognosis, such as poor histological differentiation, mucinous

differentiation, venous invasion, and a greater systemic inflammatory response (Nawa et al., 2008; Patel et al., 2018). Additionally, disease site has been shown to correlate with the presence of oncogenic mutations, with right sided disease harbouring significantly more *RAS*, *KRAS*, and *BRAF* mutations (Bylsma et al., 2020).

1.3.3 Microscopic Anatomy of the Large Bowel

The microscopic anatomy of the large bowel is organised into the same general structure as the rest of the gastrointestinal tract, an arrangement of concentric layers of tissues with individual structures and functions. Beginning at the innermost layer, immediately adjacent to the lumen is the mucosa, comprising the colonic epithelial crypts (crypts of Lieberkühn) and associated connective tissue, the lamina propria which harbours a varied population of immune cells and separates the crypts from the muscularis mucosae, the thin band of smooth muscle forming outer layer of the mucosa. Beneath the mucosa is the submucosa, a layer of loose connective tissue providing support to the mucosa during peristalsis, which is conducted by the next layer, the muscularis propria. The muscularis propria is a double layer of smooth muscle with the inner (circular) layer fibres wrapped round the long axis of the bowel and the outer (longitudinal) layer fibres running parallel to the long axis of the bowel. The muscularis propria is innervated by the myenteric (Auerbach's) plexus, which lies between the two muscle layers. The outermost layer of the

bowel is the subserosa, a layer of fibroadipose tissue that is lined by the serosa, a layer of cuboidal mesothelial cells. Detailed in Figure 1.3.

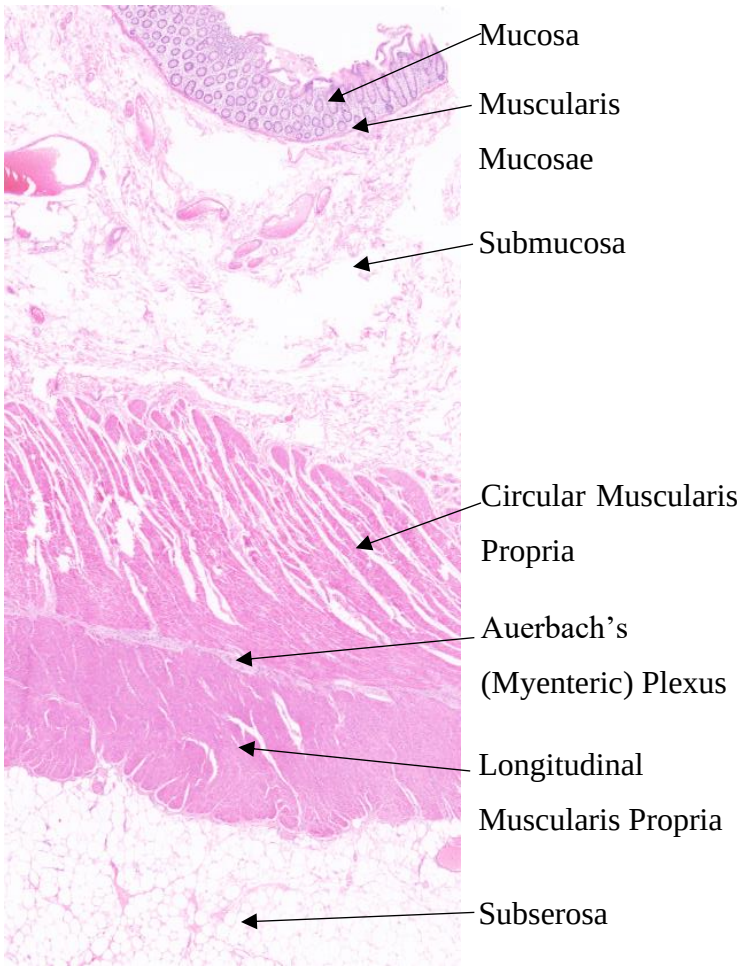


Figure 1.3 Microscopic Anatomy of the Colon. *Light micrograph of Haematoxylin and Eosin stained section of the colon, detailing key anatomical structures. NB image taken from section of T3 CRC and as such, structures may not be identical to those seen in healthy bowel.*

1.3.4 Pathological Grading

Pathological assessment of the tumour and its microenvironment is essential for predicting patient prognosis, guiding therapeutic regimens, and determining follow up to prevent recurrence. There are an ever – increasing number of pathological factors that can affect these clinical outcomes, therefore, systems that grade tumours based on the combination of multiple factors have long been seen as the clinical gold standard in pathology. Two such multifactorial systems have been implemented in routine diagnosis of CRC: Dukes' Staging Criteria and the TNM Staging Criteria.

1.3.4.1 Dukes' Staging of CRC

C. E. Dukes first described his system of classification of rectal tumours in 1932, noting that the criteria could be applied to all intestinal carcinomas, which it later was to clinically stage colonic tumours (Dukes, 1932). Based on an analysis of 215 cases, Dukes' system grouped patients into three categories according to the extent the tumour had spread through the rectal parenchyma and extra – rectal tissues. Tumours are classified Dukes' A where tumour growth is limited to the bowel wall and there is an absence of nodal involvement, Dukes' B in cases where the tumour has extended beyond the wall into the extra – rectal (or extra – colonic in the case of colon tumours) tissues but again with no involvement of regional nodes, and Dukes' C in cases where the tumour has extended beyond the bowel wall and there is involvement of regional nodes (Figure 1.4A). Three – year survival based on this system of classification was 80% for Dukes' A, 73% for Dukes' B, and 7% for Dukes' C (Dukes, 1932). A follow up study published in 1935 updated these criteria to subgroup Dukes C cases into two subdivisions based on the extent of regional node involvement (Figure 1.4B). Cases were now to be graded C_1 if the spread of the tumour was isolated to the lymph nodes below the point of ligature of the blood vessels, and C_2 if nodal involvement had reached or exceeded the point of ligature (Gabriel et al., 1935). An additional modification to these criteria was made by Turnbull and Colleagues in 1967, with the addition of a fourth group, Dukes' D, identified by the presence of distant metastases or where tumour is unresectable due to spread to adjacent organs (Turnbull et al., 1967).

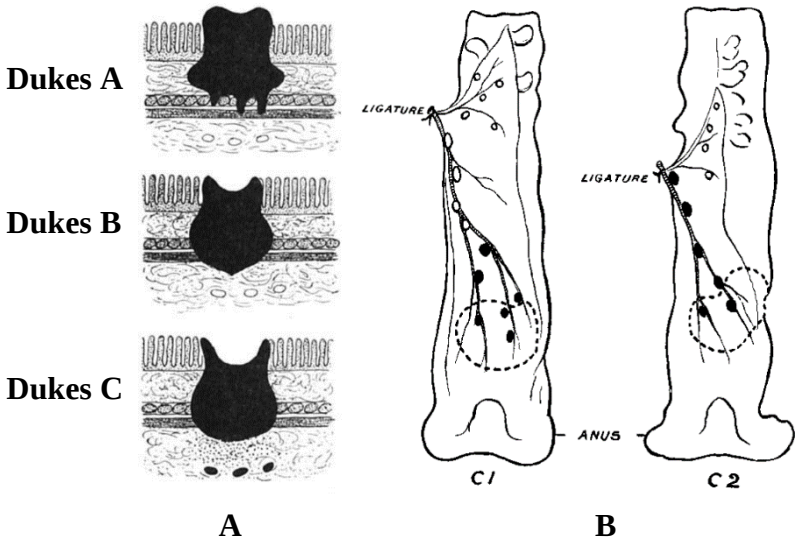


Figure 1.4 Dukes' Classification of CRC. *The classification of CRC according to Dukes' original study (A, adapted from Dukes, 1932) and assessment of nodal involvement for classifying Dukes' C_1 and C_2 (B, adapted from Gabriel et al., 1935).*

1.3.4.2 Tumour, Node, Metastasis Staging of CRC

The staging criteria that become the gold standard for CRC prognosis and has seen universal implementation is the Tumour, Node, Metastasis (TNM) staging criteria. A preliminary iteration of this criteria was originally published in 1946, however, it was not until 1987 that an internationally standardised format was determined, addressing discrepancies between the Union for International Cancer Control (UICC) and the American Joint Committee on Cancer (AJCC) versions of the criteria (Denoix, 1946; Sobin et al., 1988). The premise of this criteria is to address and subdivide each component of tumour invasion individually, with T Stage examining the extent of local tumour spread through each layer of the bowel, N Stage examining both the presence and extent of nodal involvement, and M Stage examining the extent of both local and distant metastatic spread.

The AJCC published the most recent update to the TNM criteria in 2018, and it defines T stage as such: T0 or Tis (*in situ*) – the presence of non – invasive carcinoma only in the mucosa, T1 – tumour is invasive into the submucosa, T2 – tumour is invasive into either layer of the muscularis propria, T3 – tumour is invasive into but has not extended beyond the subserosa, T4 is subdivided into two stages: T4a – tumour is invasive into the peritoneum, T4b – tumour has grown into or attached to surrounding organs. To accurately determine N stage and prevent under – staging, it is recommended that a minimum of 12 lymph nodes be excised during resection (Shia et al., 2012). If this number is achieved, then

N stage is graded as such: N0 – no presence of cancer in any lymph nodes, N1 is subdivided into three stages: N1a – presence of cancer in 1 regional node, N1b – presence of cancer in 2 / 3 nodes, N1c – no presence of cancer in regional nodes but isolated tumour foci are found in the pericolic / perirectal fat or adjacent mesentery. N2 is subdivided into 2 stages: N2a – cancer present in 4 to 6 regional nodes, N2b – cancer present in ≥7 nodes. M Stage is graded M0 if there is no evidence of distant or local metastatic spread and M1 is subdivided into 3 stages: M1a – cancer has spread to 1 distant site / organ but not into peritoneal cavity, M1b metastatic spread to 2 or more distant sites but again not into the peritoneal cavity, and M1c – metastatic spread to distant sites and into the peritoneal cavity (Weiser, 2018). In clinical practice, the T, N, and M stages are amalgamated into grouped stages that are used for prognosis and therapeutic decision making (Table 1.1).

Stage	***Dukes' Stage***	***T***	***N***	***M***	***5 – Year DFS Rate (%)***
0	*–*	*Tis / T0*	*N0*	*M0*	**100**
I	*A*	*T1 – 2*	*N0*	*M0*	**93.1**
IIa	*B*	*T3*	*N0*	*M0*	**78.3**
IIb	*B*	*T4a*	*N0*	*M0*	**73.2**
IIc	*B*	*T4b*	*N0*	*M0*	**61.3**
IIIa	*C*	*T1 – 2* *T1*	*N1 / N1c* *N2a*	*M0* *M0*	**65.4**
IIIb	*C*	*T3 – 4a* *T2 – 3* *T1 – 2*	*N1 / N1c* *N2a* *N2b*	*M0* *M0* *M0*	**56.3**
IIIc	*C*	*T4a* *T3 – 4a* *T4b*	*N2a* *N2b* *N1 – 2*	*M0* *M0* *M0*	**37**
IVa	*D*	*Any T*	*Any N*	*M1a*	**8.3**
IVb	*D*	*Any T*	*Any N*	*M1b*	**0**
IVc	*D*	*Any T*	*Any N*	*M1c*	**0**

Table 1.1 TNM Staging Criteria. *The grouped TNM staging criteria according to the AJCC 8th Edition and comparison to Dukes' stages. 5 – year DFS survival rates pertain to TNM stages* (Tong et al., 2018).

1.3.4.3 Prognosis of Stage II Disease

While the AJCC / UICC TNM staging criteria has remained the gold standard for pathological reporting and staging since its inception, it has received long standing criticism for the variable prognosis of Stage II disease, which has persisted even with further sub stratification of Stage II and Stage III. In the 8th edition of the TNM staging criteria, patients with Stage IIIa disease have a notably better 5 – year DFS prognosis than patients with Stage IIc disease (65.4% and 61.3%, respectively, Table1.1, Tong et al., 2018). This is likely due to the use of the staging criteria to guide chemotherapeutic decisions, whereby adjuvant chemotherapy is given to all patients with Stage III disease following surgical resection but the failure to distinguish Stage II patients at high risk or recurrence and death in the setting of node negative disease has led to conflicting opinions on how appropriate adjuvant chemotherapy is for these patients (Morris et al., 2007). A number of studies have now

developed methodologies based on the consideration of additional criteria to supplement TNM staging, such as prognostic pathological characteristics and measures of systemic inflammation, that are able to identify patients with high – risk Stage II disease that may benefit from adjuvant chemotherapy (Park, Watt, et al., 2016; Petersen et al., 2002).

1.3.5 High – Risk Pathological Features

1.3.5.1 Tumour Differentiation

The pathology of tumour differentiation is based on microscopic assessment of gland structure and organization within the tumour and is traditionally defined by three grades: well, moderate, and poor. Well differentiated (low grade) tumours architecturally and cytologically resemble normal or only slightly dysplastic epithelium, moderately differentiated (average grade) tumours are intermediate with more pronounced cytological and architectural abnormality than well differentiated tumours, poorly differentiated (high grade) tumours are characterized by irregularly folded, distorted, small glands (Morson & Sobin, 1976). A fourth grade of differentiation, undifferentiated, is defined by the WHO (Bosman et al., 2010a), to describe tumours of such poor differentiation that there is little morphological indication the tumour is of epithelial origin. These grades can additionally be defined by the percentage of gland formation within the tumour, with well differentiated tumours exhibiting >95% glandular structures, moderate 50 – 95% glandular structures, poor 5 – 50 % glandular structures, and undifferentiated <5% glandular structures (Bosman et al., 2010a). These morphologies are not distinct and in the majority of tumours, multiple foci of varying differentiations will be present, and the tumour is graded according to the predominating morphology. It is understood that poorly differentiated tumours confer a poorer prognosis than well and moderately differentiated tumours (Halvorsen & Seim, 1988; Marks et al., 2018).

1.3.5.2 Histological Type

Adenocarcinomas of the colon may present with uncommon histomorphological characteristics that define them as distinctly differentiated subtypes of tumour. Two of the most common of these subtypes are Signet Ring Cell Carcinoma (SRCC) and tumours with mucinous differentiation. SRCC is defined as an adenocarcinoma exhibiting signet ring cell differentiation, cells with mucinous vacuolation of the cytoplasm pressing the nucleus to the cell periphery producing a morphology akin to a signet ring, in >50% of the tumour area.

First described in 1951 by Laufman and Saphir, primary SRCC of the colon is rare, accounting for <1% of cases and is most commonly found in tumours of the stomach (Laufman & Saphir, 1951). SRCC carries a markedly poor prognosis likely due to the frequently delayed onset of symptoms, with up to 80% diagnosed at Stage III / IV (compared to 50% for conventional adenocarcinoma), although it is paradoxically diagnosed with greater frequency in patients younger than the age of 40 (Belli et al., 2014; Kang et al., 2005; Tawadros et al., 2015; Yun et al., 2017).

Carcinomas with a prominent mucinous component are subdivided based on the tumour area covered by extracellular mucinous pools; tumours with extracellular mucin covering >50% of the lesion area are designated mucinous adenocarcinoma, and those with <50% extracellular mucin by area are designated adenocarcinoma with a mucinous component (Bosman et al., 2010a). Mucinous tumours are more common than SRCC, with both subtypes accounting for 10 – 15% of CRC cases (Hugen et al., 2015) but are again generally diagnosed at a later stage, likely due to the pliability of the tumours ameliorating the severity of symptoms until the disease has progressed to a later stage. Mucinous tumours present with a different distribution of occurrence than conventional adenocarcinoma, with the majority (54.3%) forming in the proximal colon and 24.4% in the rectum whereas 30.6% of adenocarcinomas form in the proximal colon and 38% in the rectum (Hugen et al., 2013). Prognosis of mucinous tumours is variable and stage dependent; however, mucinous differentiation seems to generally confer a worse prognosis, with evidence suggesting this is due to aberrant patterns of metastasis, higher chance of recurrence from incomplete surgical resection due to the high likelihood of rupture and spillage during removal, and variable response to neoadjuvant and adjuvant chemotherapies and radiotherapy (Hugen et al., 2015; Langner et al., 2012).

1.3.5.3 Tumour Budding

The premise of tumour budding was first described in 1954 by Imai, theorizing that a sprouting pattern of tumour infiltration represented a more rapidly growing tumour (Imai, 1954). Expanding upon this, Gabbert and Colleagues studied murine colonic tumour invasion through a combination of light and electron microscopy, noting that there was a marked dissociation of the tumour morphology at the invasive front coupled with a loss of most cytological features of differentiation that potentially enabled single cells to expedite the invasive process (Gabbert et al., 1985). In modern practise, tumour budding is defined as single tumour cells or clusters of four or less tumour cells present at the invasive margin and distinct from the contiguous tumour nests. Tumour budding has been rigorously studied

and shown to be significantly prognostic marker in a wide variety of solid tumours, such as colorectal (van Wyk et al., 2015), lung (Thakur et al., 2022), breast (Gujam et al., 2015), pancreatic (O'Connor et al., 2015), endometrial (Rau et al., 2020), and head and neck cancers (Almangush et al., 2017). In the setting of CRC, budding is used to stratify patients into three prognostic groups based on assessment of a single 0.785mm^2 field of view; patients with 0 – 4 buds classified as low budding (BD1), 5 – 9 buds as intermediate budding (BD2), and ≥10 buds as high budding (BD3) (Lugli, Kirsch, Ajioka, Bosman, Cathomas, Dawson, El Zimaity, et al., 2017). Use of this criteria has repeatedly demonstrated budding to hold significant prognostic value across all stages of disease, with high budding patients possessing a significantly worse prognosis, being at greater risk of nodal metastasis in early – stage disease, and greater risk of recurrence and death in stage II disease, highlighting its potential use to identify high risk stage II patients (Bosch et al., 2013; Cappellesso et al., 2017; Lugli, Kirsch, Ajioka, Bosman, Cathomas, Dawson, El Zimaity, et al., 2017; van Wyk et al., 2019). This body of data has resulted in tumour budding being included in the WHO classification of tumours as an additional prognostic factor to be considered during the diagnostic procedure (Bosman et al., 2010b).

1.3.5.4 Venous Invasion

Venous invasion (or vascular invasion) is defined as tumour present within an endothelium lined space that is either surrounded by a rim of muscle or contains erythrocytes (Talbot et al., 1981). It is additionally advised that the presence of tumour deposits in close proximity to an artery raise suspicion of venous invasion but is only indicative and not diagnostic without identification of a venous wall (Loughrey et al., 2022). This poses a continuous problem for the use of assessment of venous invasion in routine clinical practice as it is often difficult to determine with an acceptable degree of certainty that venous invasion is present despite these criteria, and indeed, due to this venous invasion is often missed or under reported in the clinical setting, leading to extremely varied reporting of incidence ranging from 9 – 90% (Messenger et al., 2012; Stewart et al., 2007). Detection of venous invasion can be bolstered by the use of special staining, such as elastic Verhoeff – Van Gieson stain that identifies elastic fibres in the adventitia of veins and arteries but not lymphatic vessels, that have been shown to significantly improve pathologists' rate of detection and improve the prognostic significance of the assessment (Howlett et al., 2009; C. S. D. Roxburgh et al., 2010, 2014). However, the use of elastica staining is not widely implemented in routine pathology and, while venous invasion is acknowledged as a

prognostic factor, not recommended for routine use by the AJCC. Venous invasion is a significantly negative prognostic factor in all stages of CRC, having initially been examined in the extramural context, subsequent studies showed that assessment of intramural venous invasion also held prognostic significance, determining that patients with venous invasion had a lower disease specific survival and were at greater risk of metastatic spread, particularly hepatic metastases (Betge et al., 2012; Howlett et al., 2009; C. S. D. Roxburgh et al., 2010, 2014; Stewart et al., 2007; Vass et al., 2004).

1.3.5.5 Perineural Invasion

Perineural invasion is characterised by invasion of neural tissue by the tumour along the nerve sheath and is of particular interest due to the proximity of invasive carcinoma to neural structures, notably Meissner's plexus just beneath the muscularis mucosae and Auerbach's (myenteric) plexus that lies between the circular and longitudinal layers of the muscularis propria. Similar to venous invasion, perineural invasion is very difficult to detect from routine H&E sections, with reported positivity rates often contingent on the number of blocks and fields of view assessed (Ueno, Hase, et al., 2002). Additionally, no histochemical staining has been thoroughly validated to aid in the detection of perineural invasion, though some studies have reported an improvement in detection rate from 14 – 70% and 30 – 50% through assessment on sections stained for S100 through IHC (Bellis et al., 1993; Conte et al., 2020; van Wyk et al., 2017). Despite this, reporting of perineural invasion positivity appears to be more consistent than that of venous invasion, with a meta – analysis study finding that positive rate varied from 2.83 – 54.86% (Y. Yang et al., 2015). While associations of perineural invasion with disease specific survival vary between studies, the general consensus is that the presence of perineural invasion is an adverse independent prognostic factor and indicative of a more aggressive tumour phenotype, with studies noting that patients with perineural invasion have significantly worse 5 – year disease free survival and have an increased risk nodal metastasis in early stage disease (Huh et al., 2010; Ueno et al., 2013; van Wyk et al., 2017; Y. Yang et al., 2015).

1.3.5.6 Tumour Associated Stroma

The desmoplastic reaction, the activation of parenchymal fibroblasts into Cancer-Associated Fibroblasts (CAFs) resulting in upregulation of normal fibrotic wound healing responses, is a key step in the carcinogenesis of solid tumours. Under normal conditions, fibroblasts are quiescent and reside in the interstitial matrix until activation by a variety of pro-inflammatory and growth factors, primarily TGF-ß, induces production of ECM, ECM remodelling, and eventual inactivation following restoration of homeostasis (Hinz et al., 2012).

As tumours develop, paracrine signalling mediated by factors such as TGF-ß, PDGF, VEGF, IL-6, and matrix metalloproteinases (MMPs), between fibroblasts and tumour cells forms a feedback loop that results in the eventual transformation of normal fibroblasts into CAFs (Heneberg, 2016). This process occurs in three key stages; initial recruitment of normal fibroblasts to the site of the developing tumour, re-programming of the precursor cells and subsequent transdifferentiation into CAFs, and persistence of CAF properties and propagation of activated CAFs (De Wever et al., 2014). Following the recruitment and activation of CAFs, the ECM in the tumour adjacent regions undergoes remodelling mediated by CAF secretion and modification of various ECM components. The most common ECM alteration is increased secretion of fibrillar collagen, which is known to have distinctly pro-tumourigenic properties in a variety of tumour types, for example, collagen V in carcinoma of the breast (Z. H. Zhou et al., 2017). Generally, ECM remodelling and desmoplasia has tumour enhancing properties and is correlated with an increased risk of metastasis, however, certain components of the ECM are known to have contradictory effects on tumour progression. Most notably, hyaluronan (HA) exhibits both tumour suppressive and pro-tumourigenic properties dependent on the molecular weight of the HA comprising the ECM (Bohaumilitzky et al., 2017). The primary ECM component of the famously tumour-resistant naked mole rat is a high molecular weight hyaluronan, which appears to exert anti-tumourigenic effects through increased cellular sensitization to contact inhibition (Tian et al., 2015). However, in the setting of colorectal cancer, dysregulation of hyaluronan synthetase and hyaluronidase leads to the accumulation of low molecular mass HA (LMM-HA) oligosaccharides, which directly modulate pro-tumourigenic signalling and is associated with an increased risk of metastasis (Ropponen et al., 1998; Schmaus et al., 2014). In addition to alteration of its constituent components, ECM remodelling also occurs through post translational modifications such as glycosylation, degradation mediated by factors such as MMPs, and force mediated modification via integrin binding (Winkler et al., 2020).

Given that the stromal content of various tumours has demonstrated clear links with prognosis, a tissue-based method of assessing the quantity and constitution of the stroma using diagnostic H&E sections has direct clinical utility. Multiple studies investigating the stroma on H&Es have shown that simply determining the ratio of stroma to tumour area, as measured via the percentage area using 50% stroma as a cut off for high vs low risk, is able to significantly stratify patients for prognosis independent of T-Stage and other clinicopathological features (Park et al., 2014; van Pelt et al., 2018; Smit et al., 2021).

However, a 2002 study by Ueno & Colleagues examining the histological characteristics of fibrotic stroma at the invasive front of advanced rectal tumours, identified novel features pertaining to the maturity and composition of the stroma that prognostically stratified patients into three independent groups (Ueno, Jones, et al., 2002a). In this study, stroma comprised of well organised and stratified fine fibres was characterized as mature, stroma containing brightly eosinophilic hyalinized collagen bundles as seen in a keloid characterised as keloid-like or intermediate, and stroma comprising keloid-like collagen bundles surrounded by myxoid stroma as myxoid or immature, conferring good, intermediate, and poor prognosis, respectively. This method of assessment of stromal maturity was later validated for prognosis in colon cancer and demonstrated association with additional prognostic histopathological features, with more immature stroma showing significantly increased tumour budding counts and lower infiltration of lymphocytes (I. P. Nearchou, Kajiwara, et al., 2019; Ueno et al., 2004, 2017). Furthermore, studies examining nodal and hepatic metastases found a significant correlation between the stromal maturity of the primary lesion and metastatic lesion in both settings, with immature stroma perpetuating the association with poor prognosis (Ao et al., 2019, 2020).

Molecular studies examining desmoplasia maturity found that immature stroma was significantly associated with IHC-based expression of the EMT markers Zinc finger E-box binding homeobox 1 (ZEB1) and Twist-related protein 1 (TWIST1), and expression at both the mRNA and protein levels of Periostin, which is known to contribute to colorectal carcinogenesis (Hashimoto et al., 2022; Ma et al., 2020; Sueyama et al., 2021). *In Vivo* studies demonstrated that CRC cells implanted with immature stroma derived CAFs had significantly greater tumour growth and metastasis compared to those implanted with mature stroma derived CAFs. Additionally, immature CAFs expressed significantly higher levels of the secreted form of A Disintegrin And Metalloproteinases 9 (ADAM9s), and shRNA knockdown of ADAM9s abrogated the pro-tumourigenic of immature CAFs (Ao et al., 2022).

1.3.5.7 Multi – Feature Pathological Assessment

A 2002 study by Petersen and Colleagues demonstrated that combining multiple high – risk pathological features into a single score could provide significant additional prognostic value (Petersen et al., 2002). In this study peritoneal involvement, intramural and extramural venous invasion, marginal involvement, and tumour perforation were identified as independent and highly prognostic risk factors on multivariate analysis and were therefore investigated as a combinatorial assessment. Each feature was assigned a score; 1 if peritoneum involved ± ulceration, + 1 if venous invasion detected, + 1 if surgical margin involved or inflamed, + 2 if tumour has perforated bowel wall, to produce a maximum score of 5, although anything ≥ 3 is amalgamated into a single category identifying the highest risk patients. The four categories (0, 1, 2, ≥3) were each individually prognostic with 5 – year survival estimates of 94.2%, 79.5%, 54.3%, and 30.4% respectively, but were combined into a low – risk group (0/1) with a 5 – year survival of 85.7% and a high – risk group (≥2) with a 5 – year survival of 49.8%. This high vs low risk grouping was able to significantly stratify Dukes' B patients for prognosis, thus identifying high risk Stage II patients that could benefit from adjuvant chemotherapy and more rigorous surveillance.

1.3.6 The CRC Microenvironment

Early investigations into CRC pathogenesis focussed on the independent role of specific factors predominantly determined pathologically from diagnostic H&E sections. However, advancements in a plethora of techniques have revealed the symbiotic relationship that the tumour has with its microenvironment (TME) throughout the course of its development and progression. The TME is a conglomerate of extremely heterogeneous cellular types and processes that contribute to tumourigenesis through synergistic and antagonistic means. The primary component of the TME is the extracellular matrix (ECM), a collagenous matrix (desmoplastic stroma) that is deposited around the tumour following degradation of the basement membrane, which facilitates local invasion through architectural remodelling (Goetz et al., 2011). The primary cellular component of the TME is the Cancer Associated Fibroblast (CAF), activated forms of normal fibroblasts which not only maintain the ECM in response to tumour derived cytokine signalling, but also play an active role in tumour growth and invasion through secretome alterations. CAFs are known to secrete growth factors and cytokines, such as Wnt2, that promote tumourigenesis and invasion (Aizawa et al., 2019), and ultimately metastasis (Nakagawa et al., 2004). Host inflammatory cells also comprise a substantial part of the cellular component of the TME,

both those that are part of normal mucosal surveillance and those recruited to the site of carcinogenesis. The prognostic role of inflammatory cells is dependent upon the type and density of those present in the TME, some such as neutrophils are known to have broadly pro – tumourigenic effects whereas natural killer (NK) cells and the majority of lymphocytes prohibit tumour invasion, and indeed some cells such as macrophages have conflicting implications of their role in CRC progression (Flavell et al., 2010; Norton et al., 2015). The TME additionally comprises multiple other processes, such as angiogenesis (De Smedt et al., 2015; Rmali et al., 2007), and indeed organisms, derived from the gut microbiota (Wong & Yu, 2019), that play important roles in CRC development and invasion.

1.4 Subtyping of CRC

1.4.1 Omics Based Subtyping

The TNM staging criteria remains the prognostic and theragnostic gold standard in routine diagnostic pathology for CRC around the world. However, as noted in 1.3.4.3, longstanding issues in accurately determining the prognosis of patients with Stage II node negative disease and administering appropriate therapy to these patients has directed research to identify additional prognostic characteristics and develop novel prognostic classification systems than can address this unmet need. Advances in genomic and transcriptomic analysis techniques coupled with increasingly large, open – source datasets generated from clinical patient cohorts has positioned gene expression analysis at the forefront of methods being used to generate novel subtyping methodologies of CRC. Two such methodologies have recently demonstrated significantly prognostic and reproducible results for stratifying CRC patients: the Consensus Molecular Subtypes and the CRC Intrinsic Subtypes.

1.4.1.1 Consensus Molecular Subtypes of CRC

The Consensus Molecular Subtypes (CMS) were developed by Guinney and Colleagues in 2015 following the formation of an international consortium, The CRC Subtyping Consortium (CRCSC), aimed at elucidating whether decipherable subtyping patterns existed in existing gene expression datasets and analysis algorithms (Guinney et al., 2015). To generate these subtypes, the CRCSC applied six, independently developed and validated subtyping algorithms with varying numbers of subtypes (2 x 3, 3 x 5, 1 x 6) to 18 independent CRC gene expression data sets, $N = 4151$ patients, generated from a variety of gene expression platforms (Affymetrix, Agilent) and sample types (fresh – frozen, FFPE), that were uniformly pre – processed and normalized to minimize technical variation

(Budinska et al., 2013; De Sousa E Melo et al., 2013; Marisa et al., 2013; Roepman et al., 2014a; Sadanandam et al., 2013; Schlicker et al., 2012). The outputs from these subtyping methodologies assigned six individual subtype labels to each sample which were then analysed using a network – based clustering approach to identify associations between the different subtyping methodologies. The result of this was the four CMS groups (Table 1.2) designated; CMS1 – MSI Immune, CMS2 – Canonical, CMS3 – Metabolic, CMS4 – Mesenchymal, with 13% of samples not assigned to a subtype listed as non – consensus. CSM1, the immune subtype, is characterised by increased expression of genes associated with inflammatory response, MSI, MMR deficiency, high frequency of *BRAF* mutations, and hypermutated and hypermethylated status. CMS2, the canonical subtype is characterised by CIN, SCNA resulting in high copy number gains in oncogenes and copy number losses of TSGs, left sided disease, and upregulation of WNT and MYC signalling targets. CMS3, the metabolic subtype, is characterised by CIMP – low disease, activating *KRAS* mutations, and aberrant metabolism. CMS4, the mesenchymal subtype, is characterised by TGF-β activation, advanced stage at diagnosis, overexpression of genes associated with stromal infiltration, mesenchymal phenotype, and angiogenesis induction. In addition to each subtype possessing distinct molecular signatures, they are individually prognostic with CMS1 having a notably better prognosis than the other subtypes with CMS4 conferring a notably poorer prognosis (Dienstmann et al., 2017; Guinney et al., 2015). Furthermore, retrospective analysis of subtype dependent response to chemotherapy regimens in publicly available patient data sets and clinical trial data has elucidated strong evidence that response to different chemotherapy could be predicted by the CMS. Studies have postulated that CMS1 tumours, due to the strong immunogenic response associated with this subtype, are likely to respond to targeted immunotherapy (Becht et al., 2016a). CMS4 tumours have been shown to respond better to irinotecan – based chemotherapy than oxaliplatin – based chemotherapy (Okita et al., 2018). CMS2 and CMS3 patients have both been shown to respond positively to FOLFOX chemotherapy with CMS2 showing further positive outcomes when administered anti – EGFR therapies such as cetuximab (Lenz et al., 2019; Trinh et al., 2017).

CMS 1 ***MSI Immune***	***CMS 2*** ***Canonical***	***CMS 3*** ***Metabolic***	***CMS 4*** ***Mesenchymal***
14%	*37%*	*13%*	*23%*
MSI, CIMP High, Hypermutation	SCNA High	Mixed MSI Status, SCNA Low, CIMP Low	SCNA High
BRAF Mutations		*KRAS* Mutations	
Immune Infiltration and Activation	WNT and MYC Activation	Metabolic Deregulation	Stromal Infiltration, TGF-β Activation, Angiogenesis
Good Prognosis, Worse Survival After Relapse			Worse Relapse Free and Overall Survival

Table 1.2 Characteristics of the Consensus Molecular Subtypes. *Molecular and Phenotypic characteristics of the four CMS groups. Adapted from Guinney et al., 2015.*

1.4.1.2 CRC Intrinsic Subtypes

Following the development of the CMS, multiple groups independently published studies highlighting that the genetic signatures used to define adversely prognostic, stromal infiltrate – associated subtypes in multiple subtyping methodologies contained a disproportionate number of genes expressed by stromal cells, notably cancer associated fibroblasts (CAFs), which intrinsically display mesenchymal traits. It was theorized therefore, that the use of whole tumour lysates to generate the gene expression data sets that established these subtyping criteria is introducing a large amount of non – tumour material originating from the tumour microenvironment into the analyses. These non – tumour derived transcriptomic signatures displaying mesenchymal characteristics are subsequently being used to define mesenchymal subtypes while masking signatures originating from tumour cells (Calon et al., 2015; Isella et al., 2015).

To address this issue, Isella and Colleagues aimed to develop a transcriptomic – based, tumour specific subtyping methodology that removed the effect of stromal gene signatures present in previous subtyping systems (Isella et al., 2017a). This was achieved through the use of Patient Derived Xenografts (PDXs), wherein the stromal component of the primary tumour is gradually replaced with murine constituents during the xenotransplantation process (Yoshida, 2020). The transcriptomes of 515 PDX samples

generated from 244 patients were analysed using human – specific probes and a non – negative matrix factorization clustering algorithm, resulting in the five CRC Intrinsic Subtypes (CRIS) denoted A – E, generated from an optimized cluster number. The authors note that the five subtypes can be grouped into two families based on shared characteristics, with CRIS – A / B being enriched for MSI tumours, right sided disease, mucinous histology, CIMP, and hypermutator phenotype, while CRIS – C / D / E are enriched for specific focal amplifications of known oncogenes such as MYC and aberrant WNT signalling. Specifically, CRIS – A is highly enriched for both *BRAF* – mutated MSI tumours and *KRAS* – mutated MSS tumours displaying MSI – like features, and metabolic deregulation. CRIS – B is characterised by mesenchymal signatures and high TGF – β signalling conferring pro – invasive traits. CRIS – C is characterised by high EGFR pathway signalling, CRIS – D by IGF2 overexpression, and CRIS – E by frequent mutations in *KRAS* and *TP53*(Isella et al., 2017a). This subtyping methodology is an independent predictor for prognosis and similarly to the CMS, are predictive for response to anti – EGFR therapies such as cetuximab.

1.4.2 Histopathological Subtyping of CRC

The CMS and CRIS subtyping methodologies represent an important step forward in precision medicine, examining the individual's unique genetic constitution and using it to determine prognosis and guide therapeutic decision making will inherently improve survival outcomes in the future. However, they have seen little implementation clinically and face important roadblocks prohibiting their routine use in diagnostic pathology. Both subtyping methods are determined by modern transcriptomic techniques and in the case of the CRIS requires a PDX to be grown which, at present, are far too costly to implement in a diagnostic pathology lab and carry out for each CRC patient. Additionally, Isella and Colleagues note that the xenotransplantation process involved in determining the CRIS subtypes may alter or drive certain cellular processes that define some characteristics of the subtypes, meaning that translation of this data may not be an accurate representation of the tumour in situ. This lack of clinical translatability has directed research towards developing subtyping methodologies that can be determined utilising routinely produced materials, such as diagnostic H&E and IHC sections, to produce a more clinically actionable means of stratifying patients. Investigations into developing histology – based subtyping systems have produced two highly prognostic classification systems, the Glasgow Microenvironment Score and The Phenotypic Subtypes of CRC.

1.4.2.1 Glasgow Microenvironment Score

Park and Colleagues first introduced the Glasgow Microenvironment Score (GMS) in 2015 as a prognostic measure of the interaction between the tumour and the microenvironment (Park et al., 2015a). The GMS is a combinatorial score comprising two histology – based assessments of the tumour microenvironment that have been validated in a number of patient cohorts: the Klintrup – Mäkinen (KM) grading criteria of peritumoural inflammatory response and the Tumour Stroma Percentage (TSP, elsewhere in the literature referred to as the Tumour Stroma Ratio or TSR) as a measure of the relative stromal component of the tumour microenvironment.

The prognostic significance of a prominent peritumoural inflammatory response in colorectal tumours has long been understood, with studies examining this association dating back to the early 1930's (Maccarty, 1931a). Evidence of this prognostic association was strengthened by a series of studies, beginning with Spratt and Spjut in 1967, using diagnostic H&E sections to semi – quantitatively grade the intensity of both the peritumoural and intratumoural lymphocytic infiltrate, frequently coupled with additional immune cell types,

across all stages of disease and correlate this with patient outcomes, though with somewhat conflicting results (C. S. D. Roxburgh & McMillan, 2012; Spratt & Spjut, 1967; Thynne et al., 1980). This culminated in the work of Jass, who published the first iteration of a reproducible system of assessment for lymphocytic infiltrate at the invasive margin following a study identifying this response as a stage – independent, favourable prognostic factor in 447 Duke's A – C rectal tumours (Jass, 1986). A year later, Jass and Colleagues further expanded upon this assessment by incorporating the criteria into a novel prognostic classification system of rectal (and by extension colonic) tumours, that additionally assessed the involvement of tumour growth with the bowel wall, morphology of the invasive margin, and the extent of nodal involvement to produce a cumulative score that stratified patients into four independently prognostic groups (Figure 1.5, Jass et al., 1987).

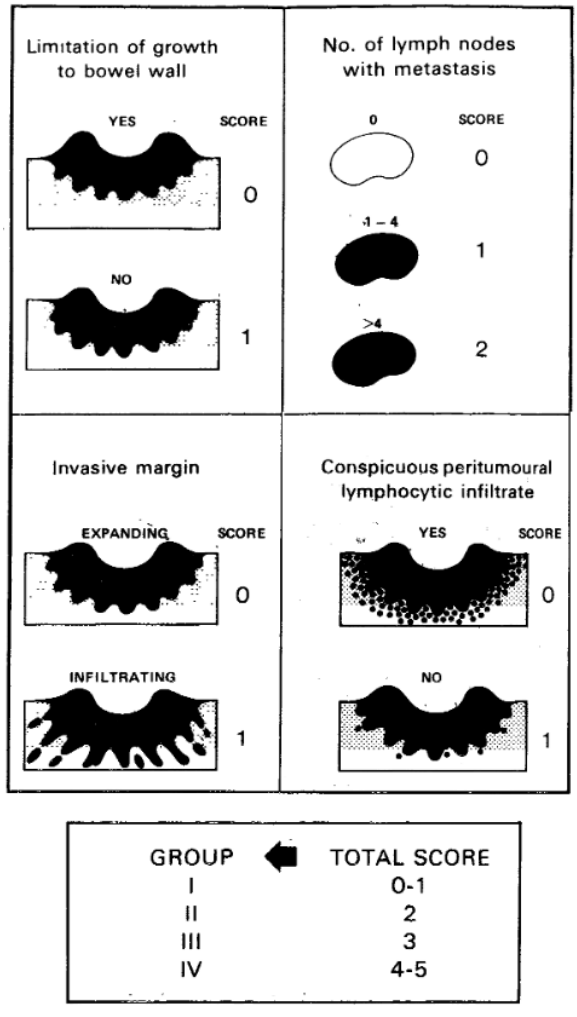


Figure 1.5 Jass Prognostic Classification System of Rectal Cancer. *Assessment criteria and scoring methodology of Jass and Colleagues novel prognostic classification system of rectal cancer. Adapted from Jass et al., 1987.*

Subsequent studies utilised advances in immunohistochemical techniques to examine the prognostic effect of specific immune cell subtypes, with a particular focus on lymphocyte subsets, and conduct more quantitative forms of assessment through positive cell counting and image analysis. The standout work resulting from these developments is the Immunoscore, developed by Galon and Colleagues in 2006, a quantitative measure of $CD3^+$ and $CD45RO^+$ lymphocyte densities both at the invasive margin and within the tumour core to produce a highly prognostic combination score able to outperform TNM staging for predicting survival outcomes (Galon et al., 2006, 2014). While assessment methodologies of this type have significant clinical and prognostic benefit, the cost associated with producing additional IHC sections for each patient coupled with the use of proprietary image analysis software has severely limited the use in routine diagnostics, and as such, more readily translatable measures of inflammatory infiltrate are required.

In 2005, Klintrup and Colleagues sought to expand further upon the work of Jass and develop a more reproducible assessment criteria of the peritumoural lymphocyte infiltrate that can be performed on clinical H&E sections. This work resulted in a four – point scoring system qualitatively describing common morphological features of the peritumoural inflammatory response, ensuring reproducibility between observers and institutions, that is independently prognostic for 5 – year survival (Klintrup et al., 2005). To stratify patients for survival, the four – point score was amalgamated into a two – point, high vs low inflammatory response score (Table 1.3) with associated 5 – year survival rates of 87.6% and 47%, respectively, and moderate to substantial interobserver agreement, with Cohen's κ scores ranging from 0.504 – 0.794, averaging 0.672.

Two – Point Scale	*Four – Point Scale*	*Description of Grade*
Low Grade Inflammation	***Absent (0)***	No increase of inflammatory cells
	Mild (1)	Mild and patchy increase of inflammatory cells at the invasive margin, but no evidence of invasion of tumour nests
High Grade Inflammation	***Moderate (2)***	Inflammatory cells form a band – like structure at the invasive margin with some destruction of tumour islands
	Strong (3)	Very prominent inflammatory reaction, forming a cup – like zone at the invasive margin, destruction of tumour islands was frequent and invariably present

Table 1.3 Assessment Criteria of Klintrup – Mäkinen Grading System. *Assessment criteria of each grade of inflammatory response according to the Klintrup – Mäkinen system and amalgamation into two – point scale. Adapted from Klintrup et al., 2005.*

This method of assessing the peritumoural inflammatory response from routinely produced diagnostic H&E sections has yet to be surpassed in terms of reproducibility and ability to predict survival outcomes, and with subsequent validation in multiple independent cohorts and additional cancer types, is the most probable candidate for implementation in diagnostic pathology (Huh et al., 2012; Mohammed et al., 2013; Richards et al., 2012, 2014; C. S. Roxburgh et al., 2013; C. S. D. Roxburgh & McMillan, 2012; C. S. D. Roxburgh et al., 2009a).

The tumour associated stroma is a prominent feature of the tumour microenvironment, comprising a variety of cellular types and dynamic processes that are known to play a vital role in the development, progression, and eventual spread of multiple solid malignancies. The predominant cell type in the stroma is activated forms of fibroblasts, known as Cancer Associated Fibroblasts (CAFs), which secrete growth factors, pro – angiogenic factors, cytokines, and metabolites which ultimately act to induce epithelial – mesenchymal transition (EMT) in tumour cells, resulting in a stem cell – like phenotype that contributes to tumour invasion, subsequent relapse, and metastasis (Calon et al., 2012a; Freeman et al., 2013; Merlos-Suárez et al., 2011; Vermeulen et al., 2010; Waghray et al., 2013). Given the number of vital roles played by the stroma in adenocarcinomas and the associated clinical implications, a means of readily assessing the potential impact of the

stroma for the individual patient would be of great clinical utility. Initial studies investigating a link between the tumour associated stroma and prognosis focussed on qualitative morphological assessment of the type and maturity of the stroma with some examining the potential impact of varying proportions of stroma at the invasive margin (Halvorsen & Seimt, 1989; Ueno et al., 2004; Ueno, Jones, et al., 2002b).

In 2007, Mesker and Colleagues conducted a study investigating the carcinoma percentage (CP), the percentage of the tissue area covered by tumour the complementary percentage of which is the stromal percentage, in 122 Stage I – III CRC patients finding that low CP was an independent adversely prognostic factor for both DFS and OS (Mesker et al., 2007). This was the first study investigating the link between prognosis and what would later become known as the Tumour Stroma Percentage (TSP) / Tumour Stroma Ratio (TSR) in CRC. Of note, is the methodology of this paper establishes robust principles of the TSP assessment that would later go on to be confirmed and repeatedly validated in multiple studies and cancer types, such as the use of the section showing the deepest point of tumour invasion, utilising the highest TSP FOV with cancer cells at all four edges to characterise the tumour, exclusion of necrosis and mucin from assessment, and the use of 50% TSP (CP in the original paper) to stratify patients for high vs low stromal component. Since this study, the TSP has been validated in various types of malignancies, such as oesophageal (Che et al., 2018), hepatocellular (Lv et al., 2015), cervical (Liu et al., 2014), breast (Kramer et al., 2019), pancreatic (Leppänen et al., 2019), and gallbladder cancer (H. Li et al., 2017) with remarkable methodological consistency across the tumour types. In the setting of CRC, the TSP has been validated as an independent, adverse prognostic factor across all stages of disease in multiple geographically independent cohorts (Hynes et al., 2017a; Mesker et al., 2009; Park et al., 2014b; van Pelt et al., 2016; West et al., 2010), has been shown to outperform tumour budding both in terms of prognostic stratification and interobserver reliability (M. A. Smit et al., 2021b), and identify high risk Stage II patients to augment the criteria used to determine chemotherapy regimens and improve survival (Huijbers et al., 2013). While features such as budding are closer to clinical implementation due to the establishment and validation of a consensus assessment criteria (Lugli et al., 2017), the breadth of prognostic and theragnostic information provided by the TSP assessment has made its clinical utility very clear, and efforts to standardize reporting for inclusion in routine diagnostics have resulted in a prospective international reproducibility study (M. Smit et al., 2019).

The combination of these two metrics as assessed through the GMS criteria (detailed in Table 1.4) provides a granular insight into the interaction of the tumour with constituents of its microenvironment, as such, efforts to validate its potential utility have yielded promising data.

	GMS 0	***GMS 1***	***GMS 2***
KM Grade 0/1 vs 2/3	High	Low	Low
TSP Status </> 50%	Any	Low	High
Cancer Specific Prognosis	Good	Intermediate	Poor

Table 1.4 Assessment Criteria of the Glasgow Microenvironment Score. *Components and scoring methodology of the GMS and associated cancer specific prognoses.*

Initial validation of the GMS in 307 Stage I – III patients demonstrated a strong association with CSS, independent of TNM stage. The GMS stratified patients into three individually prognostic groups, with GMS0 (high KM, any TSP) possessing a 5 – year survival rate of 89% whereas GMS2 (low KM, High TSP) possessed a four – fold increased risk of cancer related death and a 5 – year survival rate of 51%, and GMS1 (low KM, low TSP) a 5 – year survival rate of 75% with a near two – fold risk of death (Park et al., 2015a). Subsequent validation of the GMS in larger patient cohorts further demonstrated this prognostic capability, significantly stratifying 862 Stage I – III patients for DFS and RFS from a retrospective clinical cohort and 2912 Stage III and high – risk Stage II patients for DFS from a clinical trial derived, translational cohort (Alexander et al., 2021a). Furthermore, in the clinical trial derived cohort the GMS showed a significant association with chemotherapy regimens, demonstrating that GMS0 patients receiving FOLFOX chemotherapy had a 5 – year DFS rate of 88% whereas those receiving CAPOX possessed a 5 – year DFS rate of 62%, an effect that was not seen in the other two GMS groups. It is also of note that this association was not seen in terms of chemotherapy duration, leading the authors to conclude this effect must be of biological origin (Alexander et al., 2021).

1.4.2.2 Phenotypic Subtypes of CRC

Transcriptomics based subtyping methodologies provide remarkable insight into the molecular landscape of individual tumours, yet none of these systems have been implemented into routine clinical pathology due to the costly and time – consuming techniques used to conduct the analyses. Therefore, subtyping systems utilising routine histology offer a much more readily translatable approach to subtyping patients. To this end, Roseweir and Colleagues conducted a study looking to translate the phenotypic signatures of each of the CMS subtypes to a tissue – based assessment that may be easily conducted by a pathologist during the diagnostic process while retaining the prognostic and theragnostic properties of the system (Roseweir et al., 2017).

In CMS 1, the MSI immune subtype, Roseweir and Colleagues noted that MSI was a key feature of this subtype and one that has an extensive body of research examining its role in the development of tumours and association with multiple prognostic measures (Popat et al., 2005b). However, given that IHC for MMR protein expression is already somewhat routinely performed for certain subsets of patients to determine MSI, this would add little additional prognostic information. Furthermore, there are other transcriptomic characteristics associated with CMS 1, for example CIMP and activating *BRAF* mutations, and as such, simply testing for MSI status may not be sufficient to accurately delineate this subtype through histological means. Therefore, the relationship between MSI tumours and other phenotypic characteristics was explored to elucidate another means of identifying CMS 1 patients. It is understood that MSI tumours are associated with prognosis due to the effect of prominent local inflammatory responses, with multiple studies identifying that the presence of lymphocyte populations at the invasive margin and within the tumour core confers a significantly improved prognosis over MSS and immunogenically cold tumours and indeed that CMS 1 and CMS 4 associated with inflammatory gene expression (Becht et al., 2016b; De Smedt et al., 2015; Deschoolmeester et al., 2011; Park, Powell, et al., 2016). From this, Roseweir and Colleagues postulated that a histology – based measure of the local inflammatory response would sufficiently classify CMS 1 patients and examined the possibility of using either the Immunoscore or the Klintrup – Mäkinen system. Park and Colleagues conducted a study comparing the prognostic efficacy of both systems in the same patient cohort and found that each stratified patients for high vs low immune response to an almost identical degree, with the Immunoscore stratifying patients for CSS from 93% to 61% and the KM grade 88% to 66% (Park, McMillan, et al., 2016). Given that the use of IHC and proprietary digital pathology software is required to determine an individual's Immunoscore,

it was determined that the KM grade was the most appropriate means by which to identify CMS 1 – like patients through histopathological assessment.

In CMS 4, the mesenchymal subtype, a high number of SCNAs are the identifying feature of this subtype, specifically in genes relating to TGFβ expression, induction of angiogenesis, and increased stromal infiltrate although as previously stated, this may be due to the use of whole tumour extracts introducing a large numbers of CAF derived genes into the analyses. Nonetheless, TGFβ secretion by CAFs has been linked to adverse prognostic outcomes through promotion of metastasis and may indeed be associated with subtypes possessing stromally dense tumours (Calon et al., 2012b, 2015). Therefore, following on from work conducted prior to the establishment of the GMS identifying patients at increased risk of cancer related death and metastasis resulting from a high stromal component within the tumour microenvironment, the TSP presented a clinically validated means by which to stratify patients with a poor prognosis independent of lymphocyte infiltrate through histopathological assessment, and was identified as the optimal way to identify CMS 4 – like patients.

The transcriptomic characteristics identifying CMS 2 and CMS 3 patients proved notably difficult to translate to tissue – based analysis. As with CMS 4, CMS 2 is discernible through SCNA analysis which is only achievable through whole genome sequencing, a technique with a considerable associated time burden. Roseweir and Colleagues noted that the SCNAs primarily affected the WNT and MAPK pathways in CMS 2, alterations in which would be detectable through IHC – based expression analysis, although antibodies that could be used to evaluate components of these pathways have yet to be diagnostically validated. Aberrant WNT signalling is crucial to the development of pre – malignant lesions and subsequent carcinoma progression (Najdi et al., 2011), and is often succeeded by upregulation of nuclear localisation of β – catenin, mediated either through mutation of β – catenin itself or activating *KRAS* mutations promoting nuclear translocation (Lugli et al., 2007). A notable phenotypic characteristic associated with aberrant β – catenin function is an increased proliferative rate, positing two possible means by which to identify this subtype through tissue – based analysis: mitotic figure indices and Ki67 IHC. Though both markers provide an indication of the proliferative activity of the tumour through tissue – based assessment, it has been well documented that practically the two do not correlate, likely due to Ki67 being expressed at all stages of the cell cycle of proliferating cells while mitotic figures are only identifiable during specific stages of mitosis (Bouzubar et al., 1989), and as such, could identify different subtypes of patients. Mitotic figure indexing is conducted on

H&E sections, as opposed to Ki67 which requires IHC staining, however the method by which this is assessed is extremely time consuming, open to significant interobserver variability due to the nature of identifying mitotic figures, and in the setting of telepathology, would require high resolution scans of appropriate quality to be able to accurately identify mitoses. While Ki67 requires an additional section to be cut from the diagnostic block for IHC, it is a diagnostically validated biomarker used in routine breast cancer pathology to distinguish between Luminal A and Luminal B molecular subtypes (Goldhirsch et al., 2013), has been shown to have prognostic significance in CRC (Melling, Kowitz, Simon, et al., 2016; Reimers et al., 2014), and importantly, was identified as a marker to distinguish a CMS 2 – like subtype in one of the methodologies comprising the CMS (Roepman et al., 2014a) as well as having been shown to correlate with *KRAS* mutations in CRC (Kocián et al., 2011; Nash et al., 2010a).

It was therefore determined that Ki67 IHC posed the most readily translatable and reliably reproducible means of identifying CMS 2 – like patients in a clinical setting. CMS 3, the metabolic subtype, proved the most difficult to identify a specific biomarker to classify this subtype. *KRAS* mutations were also a common characteristic of this subtype, with ~75% of CMS 3 tumours harbouring mutations to this gene, however, given the low number of cases assigned to CMS 3, *KRAS* mutant tumours would likely align with one of the other three subtypes as a priority. Roseweir and Colleagues identified metabolic deregulation as the most prominent feature of CMS 3 tumours, however, studies associating aberrant metabolism with prognosis in CRC did so through gene expression analyses similar to that used in the development of the CMS (Vargas et al., 2014). As such, no appropriate biomarker could be identified to classify CMS 3 – like tumours through clinically translatable means and given that CMS 3 is the only subtype to show low proliferative activity, it was decided that CMS 3 – like tumours would be identified through low expression of the markers used to identify the other subtypes.

This review into histological characterization of the phenotypic signatures of the CMS resulted in the determination of the Phenotypic Subtypes of CRC, a novel classification system able to stratify patients into four, individually prognostic groups denoted Immune, Canonical, Latent, and Stromal. The criteria initially proposed to assess this subtyping method was a combination of the two – scale KM grade (1.4.2.1), which is assessed with highest priority, the TSP stratified at </> 50% stromal component and assessed with secondary priority, and the Ki67 % positivity index (or proliferation index, PI) assessed with

tertiary priority and initially stratified at </> 50% PI but later changed to </> 30% PI following subsequent analysis in later studies (detailed in Table 1.5).

Consensus Molecular Subtype	***CMS 1 MSI Immune***	***CMS 2 Canonical***	***CMS 3 Metabolic***	***CMS 4 Mesenchymal***
Phenotypic Subtype	***Immune***	***Canonical***	***Latent***	***Stromal***
KM Grade 0/1 vs 2/3	High	Low	Low	Low
TSP Status </> 50%	Any	Low	Low	High
Ki67 Proliferation Index </> 30%	Any	High	Low	Any
Cancer Specific Prognosis	Best	Good	Poor	Worst

Table 1.5 Assessment Criteria of The Phenotypic Subtypes of CRC. *Assessment criteria for KM grade, TSP, and Ki67 %PI to determine the Phenotypic Subtypes, as established by Roseweir et al., 2017. Also shown are the CMS subtypes from which the Phenotypic Subtypes are derived and their associated CSS prognoses.*

Roseweir and Colleagues conducted an initial pilot study following the translational review to investigate the association between the newly established subtypes and CSS. In a cohort comprising 237 Stage I – III CRC patients, the Phenotypic Subtypes significantly stratified patients for CSS ($P < 0.001$) with the Immune subtype showing the best prognosis, the Canonical subtype showing intermediate good prognosis, the Latent subtype showing intermediate poor prognosis, and the Stromal subtype showing the poorest prognosis for CSS. Additionally, the Phenotypic Subtypes showed a statistically significant association with OS, although to a lesser extent than with CSS ($P = 0.043$, Roseweir et al., 2017). The prognostic potential of this novel subtyping methodology was further demonstrated in a subsequent study, where the Phenotypic Subtypes significantly stratified an 893 Stage I – III patient cohort (HR 1.15, $P = 0.002$), a 146 Stage I – III patient cohort where the assessment was conducted by an external group (HR 1.41, $P = 0.006$), and a 1343 Stage II – III patient clinical trial cohort (HR 1.19, $P = 0.008$) all for DFS. Furthermore, in the clinical trial cohort, the Immune subtype showed a significant association with chemotherapy regime, whereby

patients of this subtype had a significantly improved prognosis when administered FOLFOX over CAPOX. These data demonstrate that histology – based subtyping is highly translatable as well significantly prognostic and theragnostic.

1.5 Digital Pathology

The concept of conducting pathological analysis on media other than a physical glass slide is not a novel concept. In 1986, Ronald Weinstein described Telepathology as "the practice of pathology by visualizing an indirect image on a television screen rather than viewing a specimen directly through a microscope", noting that this practice that had already been thoroughly investigated in the setting of radiography, but television monitors had yet to be designed with the appropriate number of lines of resolution to move the field of telepathology out of its infancy (Weinstein, 1986). The continuing development of telepathology in the following decades led to the eventual validation and approval by the FDA of high – resolution digital image scanners, capable of generating Whole Slide Images (WSI) at 40x objective magnification, and various slide viewing software systems to allow pathologists to view, share, and diagnostically assess routinely produced histology slides remotely. With the advent of Artificial Intelligence for image segmentation tasks, the next step for the clinical application of digital pathology is the incorporation of targeted image analysis algorithms to aid in the diagnostic workflow, a notion that has been gaining traction over recent years.

1.5.1 Traditional Image Analysis

Investigations of the potential of cellular subtyping through automated morphometry date back to the mid – twentieth century, with characteristics such as cell area and nuclear diameter being examined to identify malignant cells in cervical smears (Spriggs, 1969). This paradigm of taking qualitative features used by pathologists to identify clinically relevant morphologies and translating them to quantitative assessment by assigning measurable values forms the backbone of traditional image analysis, although this methodology is still relevant and routinely utilised today. The development of histological staining techniques such as IHC (and later immunofluorescence, IF) provided ground truth for cellular subtyping and coupled with the high throughput medium of Tissue MicroArray (TMA), prompted the development of image analysis algorithms able to generate clinically relevant data (Camp et al., 2002).

Early image analysis studies were reliant upon the use of handcrafted feature vectors and manual annotation of tissues and cells of interest to quantify meaningful morphologies on histology images (Gurcan et al., 2009), however, careful and targeted application of these

variables demonstrated the impact that image analysis could have on disease prognosis. Mitmaker and Colleagues in 1991, demonstrated that nuclear shape factor derived from colorectal histology images was able to stratify 100 Dukes' A – C patients for survival to a greater significance than Dukes' grade and histological differentiation, with a hazard ratio of 11.40 (Mitmaker et al., 1991). This study had a significant drawback however, in that generating the nuclear shape data to conduct the analysis required manual annotation of fifty interphase cells per patient, taking 20 – 30 minutes to annotate a single case, thus not delivering on the promise of image analysis to reduce the time burden of generating clinically relevant quantitative data. Subsequent developments made in automated nuclear segmentation vastly reduced the annotation requirement for image analysis studies and improved the quality of association of outputs from algorithms with clinical outcomes (Bamford & Lovell, 2001). Automated cell detection permitted investigation into the prognostic role of a variety of biomarkers based on sub cellular localization. In the setting of CRC, elevated levels of nuclear β – catenin quantified through automated assessment of IF labelled TMA sections from 310 patients, showed an association with increasingly poor prognosis where pathologist assessment failed to significantly stratify the same patients (Camp et al., 2002). A similar study in 583 CRC patients investigating the effect of MET localisation, quantified automatically on IF labelled TMA sections, identified the ratio of membranous to cytoplasmic expression as an independent prognostic factor in Stage I and II disease (Ginty et al., 2008). Interestingly, when the same study utilised IHC stained TMA sections instead of IF, no significant association was found between MET expression and prognosis, highlighting the importance of appropriate staining modalities for image analysis applications. The next task faced by image analysis in furthering its clinical utility is global tissue segmentation and classification, the ability of algorithms to automatically distinguish important tissue morphologies to further reduce the need for manual annotation of regions of interest (ROI) for analysis. Some studies utilising machine learning algorithms such as support vector machines (SVM) and random forest coupled with additional feature descriptors such as texture measurements showed promise in certain classification tasks (Kather et al., 2016; Komura & Ishikawa, 2018), however, it would not be until the widespread application of deep learning methodologies to digital pathology that reliably accurate tissue segmentation translatable across large numbers of patients became a reality.

1.5.2 Deep Learning in Pathology Image Analysis

The theory underpinning neural networks, the foundation of modern deep learning research, is not a novel concept. The studies of McCulloch and Pitts in 1943 and Rosenblatt in 1958 describe the fundamental principles of artificial neurons that are able to learn from data through self – adjusting weights and thresholds, conceptually derived from the function of biological neurons (McCulloch & Pitts, 1943; Rosenblatt, 1958). However, it would take another 50 years of theoretical and technological development to begin to realise the potential of neural networks. The performance superiority of neural networks was first actualised in 2012 with the development of AlexNet, a GPU implementation of a neural network architecture utilising convolutional functions (termed convolutional neural networks, CNNs) first described in the work of Fukushima and expanded upon with the development of LeNet (Fukushima & Miyake, 1982; Krizhevsky et al., 2012; LeCun et al., 1998). While AlexNet was not the first CNN, its architecture incorporated novel features that would become standard approaches in subsequent models, such as the use of Rectified Linear Units (ReLU) instead of the traditional tanh function which drastically reduced training time and the use of dropout, whereby the output of hidden neurons with a given probability (usually 0.5) is set to 0, which reduces model overfitting by only activating a subset of neurons that contribute positively to the classification (Hinton et al., 2012). These developments resulted in AlexNet comprehensively winning the 2012 ImageNet Large Scale Visual Recognition Challenge, an image classification competition involving 15 million images labelled with 22000 classes, with a top – 5 error rate of 15.3% compared to the second – place top – 5 error rate of 26.2%. Following the development of AlexNet, CNNs were soon being applied to histology image tasks, concurrent with the development of biomedical image specialised network architectures that were able to better handle the gigapixel WSI images, such as U-net (Ronneberger et al., 2015a). These later network architectures would eventually be included in off the shelf image analysis software packages such as Visiopharm (Watson et al., 2020)and HALO (I. P. Nearchou, Lillard, et al., 2019a), enabling tissue – based deep learning studies to be conducted without high level programming requirements in settings.

Early studies utilising CNNs in histology image analysis focussed primarily on survival prediction, slide – level classification, and computer – aided diagnosis (CAD). One of the first disease modalities thoroughly studied through deep learning was that of prostate cancer. This is likely due to the availability of a vast quantity of needle core biopsies and the fact that the Gleason grading system, used as the primary prognostic tool, is a subjective morphological pattern – based assessment with notoriously variable interobserver agreement

that would therefore lend itself well to interpretation by neural networks. Initial attempts to automate Gleason grading through deep learning yielded promising results although on limited datasets with homogenous grades (N. Zhou et al., 2017). A 2018 study by Arvaniti and Colleagues comparing the best performing network architectures of the time showed that MobileNet (Howard et al., 2017) based assessment of Gleason grading on TMA cores could accurately grade cases concurrently with specialist pathologists, identify relevant morphologies to determine grade, and prognostically stratify patients for DFS to a greater significance than pathologist's assessment (Arvaniti et al., 2018). In the field of CAD, a seminal work published by Campanella and Colleagues, a model trained and validated on 44,732 prostate, skin, and axillary lymph node slides from 15,187 patients, totalling a pixel count 88.4x that of the ImageNet dataset, without any prior curation or annotation achieved an unprecedented level of accuracy for detection of potentially malignant abnormalities, with a AUCs >0.98 for all cancer types tested (Campanella et al., 2019a). Sections were binarily labelled at the slide – level for the presence or absence of cancer for training, tiled into 224 x 224 – pixel images, and sections classified as positive if a single tile was positive for cancer, the potential clinical impact of which is a 65 – 75% reduction in the number of sections assessed by a pathologist while retaining 100% sensitivity. This study resulted in the production of the first FDA approved, deep learning – based CAD system. In the setting of breast cancer, an important prognostic assessment is that of identifying metastatic foci in axillary sentinel lymph node resections, however, this is time – consuming and difficult even for experienced pathologists. Therefore, in 2016 the CAMELYON16 competition was launched to assess the ability of deep learning models to identify metastatic breast cancer on node resection slides (Bejnordi et al., 2017). On 129 nodal WSI (38% positive), the top performing algorithm achieved an AUC = 0.994, outperforming time – constrained pathologist assessment (AUC = 0.810). Subsequent studies have shown that CAD in this task significantly improves the detection rate for lymph nodes positive for metastatic breast cancer (Steiner et al., 2018).

In the setting of CRC, initial studies involving deep learning were primarily targeted towards prediction of disease specific survival from WSI, establishing the paradigm of neural networks acting as biomarkers themselves, as opposed to assessing known biomarkers. In these types of studies, no annotation of known pathological morphologies is performed, instead, slide – level annotation such as in Campanella et al., 2019, is performed for known survival outcomes to then allow the network to extract prognostically significant morphologies without supervision. This approach has shown promise in predicting survival, with Bychkov and Colleagues showing that a dichotic digital risk score determined from

TMA cores was able to significantly stratify 420 Dukes' A – D patients, outperforming pathologist assessment of a patient being at high or low risk based on just the information available to them on the TMA core (Bychkov et al., 2018a). Similarly, Skrede and Colleagues developed a novel deep learning biomarker that significantly stratified 1122 patients for cancer specific survival and was correlated with known prognostic markers but outperformed them in terms of prognostic stratification (Skrede et al., 2020). While deep learning biomarkers have the potential to amalgamate the prognostic information from all morphologies available to them on WSI and weight their importance in a single model, theoretically providing a more accurate prognosis, the question of patient trust is consistently raised whenever new studies are published. Indeed, techniques such as gradient – weighted class activation mapping, Grad – CAM (Selvaraju et al., 2016; Yosinski et al., 2015), do exist to aid in removing the "black – box" label that biomarkers of this nature have, deep learning has yet to see implementation in clinical pathology and it is likely that the first algorithms to see use will be targeted towards known biomarkers with high interobserver variability.

To this end, more recent deep learning studies in CRC have turned to established biomarkers as inputs for model training. One biomarker that has seen a recent resurgence in interest is that of MSI. It is well known that MSI tumours possess distinct morphological characteristics that have long been used by pathologists to identify patients to put forward for molecular testing (Greenson et al., 2003; Halvarsson et al., 2008). Multiple studies have now demonstrated that deep learning models are able to predict MSI status from H&E WSI to an extremely accurate degree, with AUROC values up to 0.931, often outperforming pathologists in the same task (Kather, Pearson, et al., 2019; Yamashita et al., 2021). Studies such as these demonstrate the synergistic way in which pathologists could work with deep learning algorithms to aid in directing patients for additional molecular testing with greater confidence. As such, future work in this field is becoming more directed towards strengthening the interpretability of more novel biomarkers that have yet to see clinical implementation, such as tumour – associated stroma quantification (Geessink et al., 2019; Hacking et al., 2022), immune cell subtyping (Vayrynen et al., 2020), semantic tissue segmentation for interrogating prognostic morphologies (Graham et al., 2019; Pai et al., 2021), studying the spatial relationships between the tumour and its microenvironment (I. Nearchou et al., 2021; I. P. Nearchou, Lillard, et al., 2019b), and translating molecular subtypes to clinical specimens (Sirinukunwattana et al., 2021).

1.5.3 Translation of Digital Pathology to Clinical Practice

The use of digital pathology in the diagnostic setting offers a range of benefits to pathologists, such as decentralization, which is of particular interest in clinical trials and could aid in bridging socioeconomic gaps in countries where it is difficult to deliver timely specialist histopathological diagnosis (Pell et al., 2019). Despite initial resistance to the adoption of WSI assessment over traditional glass slides as the standard for routine diagnosis, many pathology departments worldwide are making the transition to digital pathology.

In 2013, the College of American Pathologists (CAP) produced a set of 12 guidelines for validating WSI for diagnostic purposes via assessment of interobserver agreement between glass slides and WSI, which were subsequently reaffirmed in 2022 (Evans et al., 2022; Pantanowitz L et al., 2013). These guidelines formed the basis of multiple validation studies that established non-inferiority of WSI assessment to glass slide assessment in a variety of diagnostic settings, with studies additionally noting that in cases where clinically significant discordance was observed, the root cause was easily identifiable, predictable, and avoidable (Snead et al., 2016; Thrall et al., 2015). While some data has shown that WSI assessment incurs an increased time burden, it does not significantly impact turnaround times and is indeed ameliorated following training and adjustment (Hanna et al., 2019; Mills et al., 2018). In the UK, this body of evidence demonstrating non-inferiority of pathological reporting from WSI in the clinical setting led to the Royal College of Pathologists (RCPath) issuing best practice recommendations for the implementation of digital pathology, resulting in the widespread adoption of digital pathology as the standard for primary diagnosis in pathology departments (Cross et al., 2018). Indeed, this move towards primary reporting from WSI and decentralization paid dividends during the COVID-19 pandemic, where the necessity of remote work prompted an increased uptake of digital pathology for primary diagnosis within pathology departments, as well as for pathology teaching in academic departments. The use of digital pathology, coupled with appropriate guidelines for remote work, during this time facilitated continuing diagnostic practice, while easing workforce crises and mitigating the impact to reporting turnaround times (Browning et al., 2021; Williams et al., 2020).

While adoption of digital pathology for primary diagnosis has become more commonplace in recent years, the use of image analysis algorithms in the diagnostic setting has yet to see widespread implementation. By removing the need for pathologists to labouriously and time consumingly count hundreds of positive cells from IHC slides, an image analysis algorithm using a standardized staining and assessment protocol would be of

considerable benefit to already stretched pathology departments. Such algorithms, and indeed the regulatory approval thereof, are not a recent invention. In breast cancer, 501(k) clearance from the US FDA for image analysis algorithms able to clinically analyse the five key IHC based markers for molecular subtyping; ER, PR, HER2, Ki67, and p53, was granted to Roche Ventana a decade ago (Nassar et al., 2011; Welsh et al., 2013). Since then, subsequent image analysis algorithms have also gained FDA approval for clinical use with HER2 scoring after demonstrating non-inferiority to manual assessment and have been accompanied by best practice guidelines from the College of American Pathologists (CAP) (Bui et al., 2019; Qaiser et al., 2018; Trahearn et al., 2017). Furthermore, in addition to reducing equivocal HER2 cases, the HER2-Connect image analysis algorithm produced by Visiopharm has demonstrated accurate and reliable assessment of HER2 IHC in both breast and gastro-oesophageal adenocarcinomas, highlighting the potential translatability of such algorithms to a range of clinical applications (Brügmann et al., 2012; Koopman et al., 2018). Despite the fact that there are now more than 30 image analysis algorithms with regulatory approval and best practice guidelines for their implementation, validation, and maintenance, their use is yet to become routine and is often at the discretion of the pathologist and pathology department (Lara et al., 2021).

The advent of AI has the potential to expedite the adoption of image analysis in clinical practice and transform the diagnostic histopathology workflow. The most imminent application of AI to diagnostic pathology is the refinement of existing image analysis algorithms to assuage the tedium of counting individual cells and meaningfully link the outputs of these algorithms to clinical decision making. However, it is the image rich nature of pathology as a specialty that is of particular interest to those looking to implement AI in clinical practice. The move towards a fully digitised diagnostic workflow over the past decade has created a vast repository of training data for AI-based computer vision models, laying the foundation for pathology to become an AI integrated specialty. This is clearly evidenced by the surge in publications combining AI with pathology with 3398 studies being published between 2010-2020, over 500 more than radiology, the next most frequently published specialty looking to integrate AI (Meskó & Görög, 2020).

Although interest in AI within the clinical setting is ever increasing, there are still significant barriers, both regulatory and in terms of infrastructure, to making it part of the routine diagnostic workflow. There are currently few, if any, protocols in place for the initial and continuing validation of any clinically implemented AI system, validation of any training and reference datasets notwithstanding, but steps are being taken to address this. In

2018, the UK Research and Innovation (UKRI) agency awarded £50 million through the Innovate UK scheme to create 5 centres of excellence for Digital Pathology and clinical AI (GOV.UK, 2018). The aim of these centres is to establish the infrastructure and regulatory framework to enable the NHS to capitalise on advances in AI technology while ensuring its safe and effective implementation. These centres are set up to address individual components of the implementation process, for example, the Industrial Centre for AI Research in Digital Diagnostics (I-CAIRD) based in Scotland is focussed on using a collaborative network of clinicians, academics, and industry partners to identify clinical questions that would benefit from AI and develop the means by which to address them. While PathLAKE, based at the University Hospitals Coventry and Warwickshire NHS Trust, is focussed on the systems infrastructure to create a depository for the vast quantities of data required for clinical facing AI, while ensuring adherence to patient data confidentiality regulations and securities. Additionally, in 2019, the National Cancer Research Institute (NCRI) Cellular & Molecular Pathology Initiative (CM-PATH) organised a joint conference with the British In Vitro Diagnostics Association (BIVDA) to understand the current landscape of AI-based pathology tools and establish guidelines for all stages of the development and deployment lifecycle of such tools within clinical practice (Colling et al., 2019).

While opinions on when and how it will be applied within clinical practice are still mixed, it is clear that AI will inevitably play a role in the future of medicine, with pathology appearing to be the frontrunner as the first specialty to undergo revolution. The consensus amongst subject matter experts is, that by 2030, AI will increase the accuracy and aid in the standardization of diagnoses, improve detection of rare events, make assessments more quantitative, and improve the completeness and quality of pathological reporting, all of which will fundamentally bolster the quality of treatment that patients receive (Berbís et al., 2023).

1.6 Aims

This thesis aims to investigate the application of various image analysis approaches to novel histological biomarkers associated with prognosis in CRC. While the TNM staging criteria remains the gold standard for clinical prognosis of CRC, research is continually identifying novel features of the CRC microenvironment with prognostic and theragnostic significance that can be assessed from routinely produced, diagnostic H&E sections or with supplementary IHC. However, these microenvironment features have seen little clinical implementation due to variable interobserver agreement arising from non – standardised assessment criteria. It would therefore benefit the clinical application of these assessments to be conducted through image analysis, to ensure their reliability and reproducibility. To investigate potential of using image analysis for this task, the objectives of this study are as follows:

1. Develop image analysis methodologies to translate three novel, histological biomarkers to quantitative, digital pathology – based assessment: the Klintrup – Mäkinen grading criteria of peritumoural inflammatory response, the Tumour – Stroma Percentage assessment of stromal density, and the Ki67 Proliferative Index.
2. Validate digital pathology assessments for prognostic significance and determine the translatability of the algorithms across multiple, independent patient cohorts.
3. Investigate the extent to which image analysis can faithfully recreate the manual histopathological assessment and statistically analyse the agreement between the two.
4. Investigate the use and prognostic significance of applying combinations of multiple image analysis algorithms to stratify patients according to novel CRC classification system criteria.

2. Materials and Methods

2.1 Patient Cohorts

2.1.1 Norway Cohort

The Norway Cohort consisted of 299 Stage II -III CRC patients undergoing potentially curative resection between 2005 and 2015 at Kristiansand, Arendal, Flekkefjord, and Oslo Hospitals. Tumours were staged using the sixth edition of the AJCC / UICC TNM Staging criteria, and clinicopathological data was obtained from pathology reports. Follow up data was last collected in 2017, and at this time 25 (8.4%) patients had died from primary disease, 51 (17.1%) patients had died from other causes, and 222 (74.3%) patients were still alive of which 8 (2.7%) had recurrent disease. Cancer specific survival (CSS) as measured from date of surgery to date of death from cancer was used as the primary survival outcome and the mean survival time was 45 months. (Park et al., 2020).

2.1.2 Glasgow Development Cohort

The Glasgow Development Cohort consisted of 272 stage II – III CRC patients undergoing potentially curative resection between 1997 – 2007 at the Glasgow Royal Infirmary Hospital. Tumours were staged using the fifth edition of the AJCC / UICC TNM Staging criteria, and clinicopathological data was taken from the pathology reports issued following resection. Follow up data was last collected in 2013, and at this time 95 (34.9%) patients had died of primary disease, 68 (25.0%) had died of other causes, and 109 (40.1%) were still alive. CSS as measured from the date of surgery to the date of death from cancer was used as the primary survival outcome and the mean survival time was 87 months. Safehaven Number: GSH / 18 / ON / 007. (Park et al., 2015a)

2.1.3 Glasgow Validation Cohort

The Glasgow Validation Cohort consisted of 758 stage I – IV CRC patients undergoing potentially curative resection between 2000 – 2007 at the Glasgow Western Infirmary and Stobhill Hospitals. Tumours were staged using the fifth edition of the AJCC / UICC TNM staging criteria and clinicopathological data was taken from the pathology reports issued following resection. Follow up data was last collected in 2017 and at this time 229 (30.2%) of patients had died of primary disease, 264 (34.8%) had died of other causes, and 246 (32.5%) were still alive. Follow up data was missing for 19 (2.5%) of patients. CSS as measured from the date of surgery to the date of death from cancer was used as the primary survival outcome and the mean survival time was 83 months. Safehaven Number: GSH / 18 / ON / 007. (Park et al., 2015a).

2.1.4 Glasgow Screening Cohort

The Glasgow Screening Cohort consisted of 159 T Stage I – II CRC patients who underwent surgical resection following screen-detection by faecal occult blood test (FOBT) under the Greater Glasgow and Clyde (GGC) NHS board between 2009 – 2011. Tumours were staged using the seventh edition of the AJCC / UICC TNM staging criteria and clinicopathological data was taken from the pathology reports issued following resection. Follow up data was last collected in 2018 and at this time 16 (10.1%) patients had died of primary disease, 13 (8.2%) had died of other causes. CSS as measured from date of surgery to the date of death from cancer was used as the primary survival outcome and the mean survival time was 85 months. (Mansouri et al., 2016).

2.1.5 Glasgow Royal Infirmary Cohort

The Glasgow Royal Infirmary (GRI) Cohort consisted of 787 Stage I – III CRC patients undergoing potentially curative resection at the Glasgow Royal Infirmary Hospital between 1997 – 2012. Tumours were staged according to the fifth or sixth edition of the AJCC / UICC TNM staging criteria and clinicopathological data was taken from the pathology reports issued following resection. Patients were followed up for at least 3 years and the last follow up data was recorded in 2020, at this time 231 (29.4%) patients had died of primary disease, 277 (35.2%) had died of other causes, and 277 (35.2) were still alive, with survival data missing for 2 patients. CSS as measured from date of surgery to the date of death from cancer was used as the primary survival outcome and the mean survival time was 93 months. Safehaven Number: GSH / 21 / ON / 009. (Alexander et al., 2021b)

2.1.6 TransSCOT Clinical Trial Cohort

The TransSCOT Clinical Trial Cohort consisted of 2913 high risk Stage II or Stage III patients from the SCOT international, randomised, phase 3, non-inferiority trial, who underwent surgical resection between 2008 – 2013 within the UK. Patients were staged using the seventh edition of the AJCC/UICC TNM Staging criteria. Patients received either 3 or 6 months of adjuvant chemotherapy via random allocation. Chemotherapy given was either a CAPOX (capcitabine/oxaliplatin) or FOLFOX (bolus with 5-flurouracil with oxaliplatin) regimen determined by the clinician. Patients were followed up for at least 3 years and at last follow up, with disease-free survival as measured from the date of randomization to date of recurrence at any location or death from any cause being the primary survival outcome. At last follow up, 755 (25.9%) patients had recurrence / died, 2157 (74.0%) patients were still alive, and follow up data was missing for 1 patient, with the mean survival time being 35 months. NHSGGC Biorepository 16/WS/0207. (Roseweir et al., 2018a)

2.2 Histochemical Staining

2.2.1 Haematoxylin and Eosin Staining

2.2.1.1 Dewaxing and Rehydration

Formalin Fixed Paraffin Embedded (FFPE) CRC tissue sections were deparaffinized in Histoclear, then rehydrated through a descending alcohol gradient consisting of absolute EtOH, 95% EtOH, 90% EtOH, 80% EtOH, 70% EtOH, 50% EtOH, 30% EtOH, and washed in deionized water.

2.2.1.2 Staining and Mounting

Sections were stained in Harris Haematoxylin for 3 minutes and rinsed initially in deionized water, and then washed in running tap water. Sections were then dipped in acid alcohol (3% HCl in 70% EtOH), rinsed in tap water, and then in deionized water. Sections were then stained in Eosin and dehydrated in an ascending alcohol gradient consisting of: 30% EtOH, 50% EtOH, 70% EtOH, 80% EtOH, 90% EtOH, 95% EtOH and absolute EtOH. Sections were then placed in Histoclear and coverslipped using Omnimount Histological Mounting Medium (National Diagnostics, Atlanta, GA, USA).

2.2.2 Ki67 Staining

2.2.2.1 Dewaxing and Rehydration

FFPE CRC sections were dewaxed in Histoclear and rehydrated through a descending alcohol gradient of absolute EtOH, 90% EtOH, 70% EtOH. Sections were then washed in running tap water.

2.2.2.2 Antigen Retrieval

Antigen retrieval was conducted via heat induced epitope retrieval (HIER) in pH8 Tris-EDTA buffer. HIER buffer was pre-heated prior to addition of the sections, brought to pressure with sections added, and heated under pressure. HIER vessel was depressurised, allowed to cool, and the sections were washed in running water.

2.2.2.3 Blocking Procedures

Sections were then blocked for endogenous peroxidase activity in 3% H_2O_2 and rinsed in running tap water. Sections were then blocked for non-specific antibody binding using a 10% Casein solution made up in antibody dilution buffer at room temperature, after which the blocking solution was blotted from the slides.

2.2.2.4 Antibody Conditions

The sections were incubated with the mouse anti-human Ki67 primary antibody (Dako, Agilent Technologies, Cheadle, UK) at 4°C overnight. Following primary incubation, sections were washed in Tris Buffered Saline (TBS), then incubated with ImmPRESS anti-rabbit/mouse secondary antibody (MP-7500, Vector Laboratories, Burlingame, CA, USA) at room temperature, and washed in TBS.

2.2.2.5 Staining and Counter Staining

Sections were stained using DAB chromogenic substrate (SK4001, Vector Laboratories, Burlingame, CA, USA) until brown precipitate formed, then washed in running tap water. Sections were counterstained in Harris haematoxylin, rinsed in running tap water, dipped in 1% acid alcohol (HCl in EtOH), and blued in Scott's Tap Water Substitute (80mM Magnesium Sulphate, 40mM Sodium Hydrocarbonate in distilled water), before rinsing in running tap water.

2.2.2.6 Dehydration and Mounting

Sections were dehydrated in an alcohol gradient of 70% EtOH, 90% EtOH, absolute EtOH, cleared in Histoclear, and coverslipped using Omnimount Histological Mounting Medium (National Diagnostics, Atlanata, GA, USA).

2.3 Histopathological Assessment

2.3.1 Klintrup-Makinen Grading

The Klintrup-Makinen (KM) grade was used as the assessment of the local inflammatory response at the invasive margin of the tumour taken from the deepest point of invasion. Briefly, grading was conducted using a four-point scale at the invasive margin as follows(Klintrup et al., 2005): 0 – no evidence of increase in inflammatory cells, 1 – patchy immune response along invasive margin, 2 – inflammatory response forming a band like structure across a significant portion of the invasive margin, 3 – florid cup-like infiltrate at margin with evidence of tumour island destruction. Subsequently, the four-point score was amalgamated into weak or strong immune responses with grades 0-1 being classed as "weak" immune response and grades 2-3 classed as "strong" immune response.

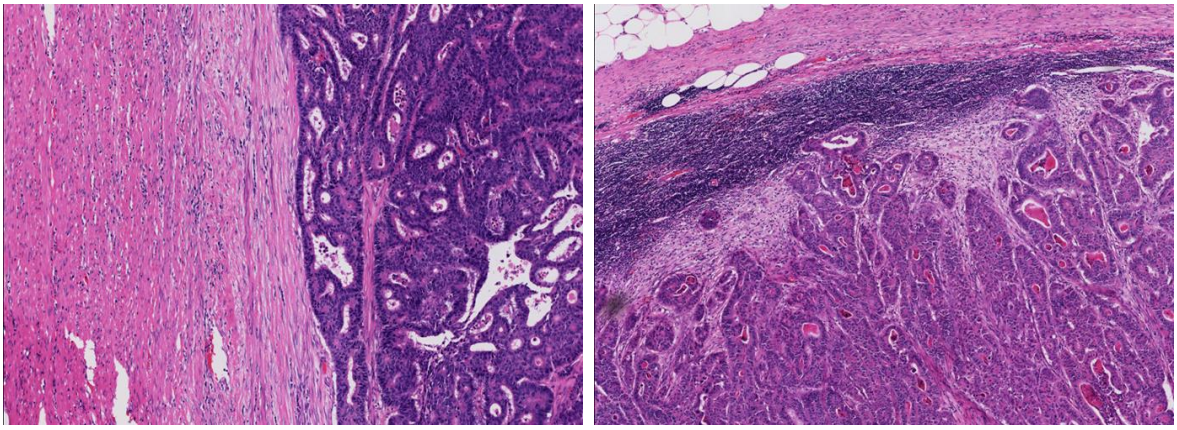


Figure 2.1 Klintrup-Makinen Grading. *H&E sections demonstrating low (left) vs high (right) KM grade. High KM grades show a band or florid cup like structure of lymphocytes at the invasive margin of the tumour and confer a better prognosis, low KM grades show little to no evidence of lymphocyte infiltrate.*

2.3.2 Tumour – Stroma Percentage

The relationship between the tumour and associated desmoplastic stroma was assessed via the Tumour-Stroma Percentage (TSP) on whole H&E sections from the deepest point of invasion as previously described(Park et al., 2014b). A representative intratumoural area was selected and assessed for the percent of the area covered by desmoplastic stroma, excluding areas of necrosis, in 5% increments. Patients were graded as TSP-Low if the stromal area was <50% and TSP-stroma if the stromal area was >50% (Park et al., 2014b).

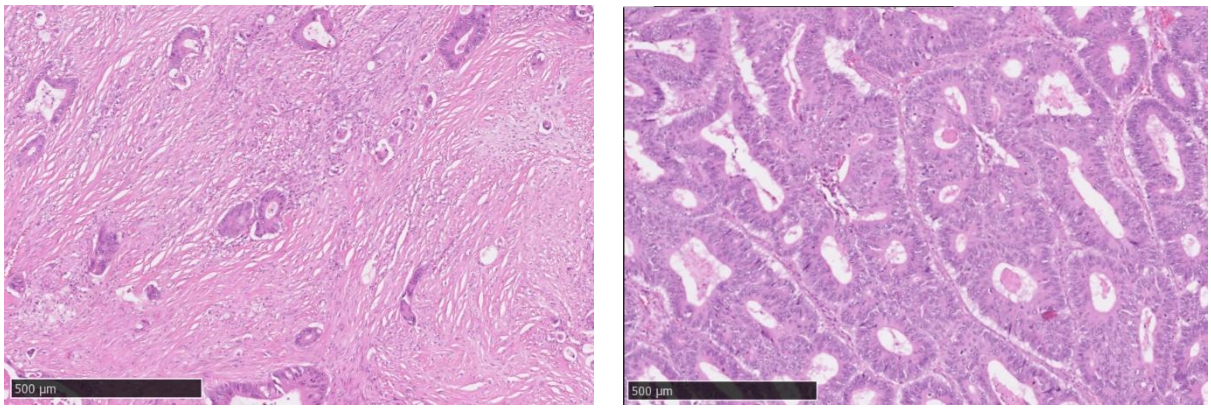


Figure 2.2 Tumour Stroma Percentage. *H&E sections demonstrating high (> 50%, left) TSP and low (<50%, right) TSP. Patients with stromal dense tumours have a worse disease specific prognosis.*

2.3.3 Proliferation Assessment via Ki67 IHC

Assessment of the proliferative capacity of the tumour was determined using quantitative assessment of Ki67, a marker expressed only in proliferating cells, via % positivity on IHC stained sections (detailed in 2.2). On full sections, this was assessed by selecting 3 fields of view within the tumour core, counting 100 cells at random and averaging the % positivity across the three areas. On TMAs, the same assessment was carried out but with one area assessed per core, if multiple cores per patient were present, then the scores were averaged across the cores per patient.

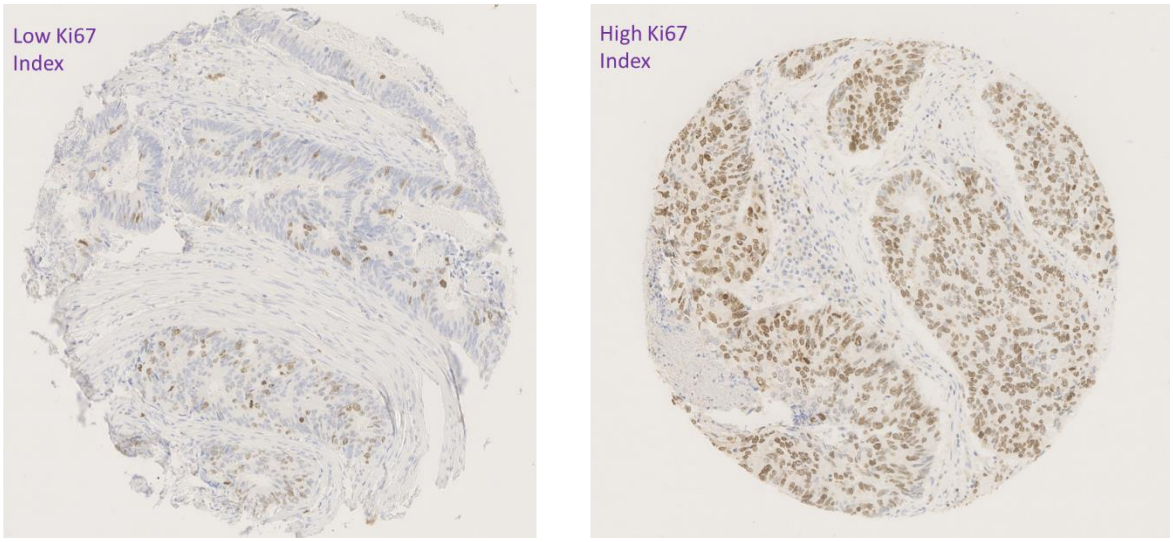


Figure 2.3 Ki67 Positivity Index. *Tissue Micro-Array cores stained via IHC showing low (<30%, left) and high (>30%, right) Ki67 %Positivity Index. In the context of CRC, patients with low Ki67 %PI have a greater disease specific prognosis.*

2.3.4 Glasgow Microenvironment Score

The relationship between the tumour, local inflammatory response, and stromal component was assessed via the Glasgow Microenvironment Score (GMS) on H&E – stained sections as previously described (Park et al., 2015a). Briefly, sections were initially assessed for the KM Grade, and patients with high local inflammatory response were assigned to GMS0, then patients with low local inflammatory response were assessed for the TSP and stratified so that high-stroma patients were assigned GMS2, and low stroma patients were assigned GMS1.

	GMS 0	***GMS 1***	***GMS 2***
KM Grade 0/1 vs 2/3	High	Low	Low
TSP Status </> 50%	Any	Low	High
Cancer Specific Prognosis	Good	Intermediate	Poor

Table 2.1 Glasgow Microenvironment Score. *Categorization of each GMS group in relation to the KM Grade and TSP Score. Patients with high (2/3) KM grade are assigned GMS0 with priority, patients with high (>50%) TSP are then assigned GMS2 and patients low for both scores are assigned GMS1.*

2.3.5 Colorectal Cancer Phenotypic Subtypes

The relationship between the tumour, local inflammatory response, stromal component, and proliferative activity was assessed via the Phenotypic Subtypes as previously described by Roseweir et al(Roseweir et al., 2017) and outlined in table 2.2. Patients were first assessed for the KM grade and those with a high local inflammatory response were assigned to the Immune subtype. Patients with a low KM grade were then assessed for the TSP and those with a high TSP were assigned to the Stromal subtype, with the remaining patients being assigned to the Canonical subtype if Ki67% positivity was >30%, and the Latent subtype if Ki67 positivity was <30%.

	Phenotypic Subtype			
	Immune	***Canonical***	***Latent***	***Stromal***
KM Grade 0/1 vs 2/3	High	Low	Low	Low
TSP Status </> 50%	Any	Low	Low	High
Ki67 Proliferation Index </> 30%	Any	High	Low	Any
Cancer Specific Prognosis	Best	Good	Poor	Worst

Table 2.2 Phenotypic Subtypes. *Categorization of each CRC Phenotypic Subtype in relation to the KM Grade, TSP, and Ki67PI. KM grade is assessed with priority, followed by TSP, then Ki67PI.*

2.4 Image Processing

2.4.1 Software

All image processing, including section annotation for training, tissue segmentation, and calculation of output variable measures pertaining to the histological assessments was conducted using the Visiopharm Oncotopix Software Version 2020.06.0.7872 (Visiopharm, Hoersholm, Denmark).

2.4.2 Training Data

Three sets of image training data for tissue segmentation were created over the course of the studies conducted for this work; two H&E based and one IHC based.

2.4.2.1 Training Data Set 1

Training Data Set 1 was created from sixteen WSI from the Norway cohort of patients and annotated manually by a single observer. To ensure as much histological variation was accounted for in the training data, slides were selected based on the patients' pathological report, ensuring that all T-Stages, differentiations including mucinous, stromal and lymphocyte densities, necrotic component, and stain intensity variation were represented. Annotations were created for 9 Classes; Tumour, Stroma, Lymphoid, Normal, Mucin, Muscle, Adipose, Necrosis, and Background, with particular attention drawn to the spatial relationship between classes to improve boundary detection.

2.4.2.2 Training Data Set 2

Training Data Set 2 was created from the publicly available CRC-VAL-HE-7K image database, originally created as a validation set for the NCT-CRC-HE-100K image database, it was used for training in this study due to computational limitations (Kather, Krisam, et al., 2019). This dataset consists of 7180 224 x 224 pixel at 0.5 microns per pixel image patches selected from N = 50 CRC patients, unevenly distributed between the same 9 tissue classes used in Training Data Set 1, with each image patch having undergone Macenko's colour-normalization(Macenko et al., 2009). Prior to training, all pixels were assigned the patch class for semantic segmentation output and all whitespace (Eosin colour deconvolved channel > 200) from each image patch was removed to reduce computation time and improve boundary detection.

2.4.2.3 Training Data Set 3

Training Data Set 3 was used exclusively for IHC with Haematoxylin and DAB segmentation tasks. This data set was constructed by a single observer exhaustively annotating 55 TMA cores and associated slide whitespace from an independent CRC patient

cohort TMA section stained with Ki67 IHC, for Tumour, Stroma, Necrosis, and Background classes, ensuring a close to even distribution of annotations of all classes on intensely and weakly stained cores to reduce network reliance on stain vectors and intensity for tissue classification.

2.4.3 Tissue Isolation

2.4.3.1 Background and Processing Artefacts

Variations in background and artefacts in tissue sections can cause stochastic outcomes during analysis processes, therefore, an initial image processing step was performed in all cases to remove scanning whitespace i.e., glass slide with no tissue section, and histological processing artefacts from the area of analysis. This was achieved by creating a feature, in RGB colour space, by subtracting the Green intensity values from the Red intensity values (i.e. Contrast Red-Green) to remove ink marks and debris. The resultant grey level image was smoothed using a stringent 53 x 53 kernel median filter, with resultant background being removed at a value of less than 10 (original scale 0-255). Post processing was used to remove small areas $<1x10^7 \mu m^2$ of tissue artefacts not contiguous with the main resection tissue followed by a hole filling process to produce a single, continuous outline of the tissue for analysis.

2.4.3.2 Lymph Node Isolation

During initial testing, it was found that the presence of lymph nodes resected with the tissue in the images was a common area of misclassification during image processing. Therefore, to remove the nodes from the analysis area, a grey level scale feature was produced based on hue and saturation of the blue staining (B' in (1)). By selecting the lesser of the two values for B' in combination with a Contrast Red-Blue feature to determine areas of high blue intensity coupled with low red intensity this enabled the haematoxylin-stained nuclei to be enhanced in relation to eosin staining.

(1) $B' = \frac{(1-r-g)G}{g}$ where $r = \frac{R}{R+G+B}$ and $g = \frac{G}{R+G+B}$ in RGB colour space

Following this step, a 55 x 55 kernel median filter and 21 x 21 kernel standard deviation filter were applied to smooth the resultant feature and improve boundary detection, respectively. Post processing involving circularity and size constraints; $\frac{4 \cdot \pi \cdot Area}{Perimeter^2} > 0.5$, area $> 3x10^6 \mu m^2$, followed by hole filling was used to selectively discard lymph nodes in

adipose tissue whilst retaining tertiary lymphoid structures closer to the tumour necessary for subsequent analysis.

2.4.4 Tissue Segmentation using Machine Learning

2.4.4.1 Training Parameters and Features

A decision forest algorithm was trained utilising 80 decision trees to a max depth of 16 per tree, with an input image of 2048 x 2048 pixels fields of view at 10x magnification. The algorithm was trained on 3 main grey level features that included combinations and transformations from 19 sub features generated from Haematoxylin and Eosin colour deconvolutions. The first feature (F1) was based on an initial haematoxylin colour deconvolution layer and utilised red-blue contrast and B' of (1) to enhance the blue values within the haematoxylin against the eosin colour deconvolution. The second feature (F2) was based on the eosin colour deconvolution and utilised contrast transformations against the haematoxylin to remove blue impurities remaining after colour separation. This enhanced differences in grey level values between the tumour and stroma. The third feature (F3) removed a combination of red-green and blue-green spectral impurities in F2, to further reduce any overlap in grey level values to distinguish between the Tumour and Stroma regions. Finally, a smoothening filter over a range of kernel sizes was applied to each feature (F1 & F3 = mean 11 x 11, F2 = median 5 x 5) to reduce the noise arising from single pixel anomalies.

2.4.4.2 Post Classification Processing

Post processing steps were used to remove areas of small, misclassified tissue and generate more contiguous classes for further analysis by converting them to the class with which they shared the longest neighbouring interface length. In addition, during initial reviews of analysis outputs areas of heavily stained stroma, misclassified as tumour, were identified. To correct for this, a new feature was generated using a grey level normalisation factor against the darkest 10% pixels of the tissue within the analysis area. This allowed for misclassified Stroma to be identified and converted to the correct class.

2.4.5 Tissue Segmentation using Deep Learning

2.4.5.1 Network Architecture and Training Parameters

Tissue segmentation using deep learning was conducted using a U-NET(Ronneberger et al., 2015b) Convolutional Neural Network (CNN) with pre-determined weights generated from network pre-training on the ImageNet dataset(Deng et al., 2010). The UNET network was further trained on study specific images in RGB colour space in 512 x 512 pixel fields of view at scanning resolution (20x) with a learning rate of $1x10^{-4}$

with Adam Optimization(Kingma & Lei Ba, n.d.) and mini-batch size 8. Training length was determined by monitoring cross entropy + Intersection Over Union (IOU) loss function until a consistently low (<0.5) rate was achieved, meaning training length ranged from ~100,000 to ~300,000 iterations dependent upon the task, in order to achieve optimal network performance. For all segmentation tasks, random rotations (90°, 180°, 270°), vertical and horizontal flipping, and brightness and contrast perturbation operations were carried out with a probability set at 0.5 for each operation. For H&E based segmentation tasks, images underwent additional hue and saturation, and Haematoxylin and Eosin colour deconvolution vector perturbation operations again each with probability 0.5. For H-DAB based segmentation tasks, the same hue, saturation, and Haematoxylin and DAB stain vector perturbation operations were performed with the same probabilities (Bándi et al., 2019).

2.4.5.2 Post Classification Processing

As all H&E and H-DAB segmentations underwent subsequent, computationally intensive image class object conversions to Regions Of Interest (ROI) to enable further analysis, pre- and post-classification operations were performed to reduce the computational burden and running times. Pixel wise tissue segmentation was performed with a classification probability of 100 / N training classes meaning each pixel was assigned the tissue class of highest probability so no pixel was left unclassified, as unclassified pixels would create a substantial number of empty ROI, which exponentially increases times required for subsequent analysis. A post segmentation class simplification operation was performed by converting clusters of pixels belonging to the same class with area <1000μm^2 to a new intermediate class and converting this class to the tissue class with which it shared the largest neighbouring boundary interface length.

2.4.6 Cellular Detection Algorithms

2.4.6.1 Lymphocyte Detection

Lymphocytes were detected from H&E-stained sections through a stepwise, threshold based image processing algorithm that initially detected all nuclei, and then selectively removed nuclei based on morphology and colour derived criteria. Firstly, a second order 5 x 5 / 11 x 11 Laplacian of Gaussian filter at 20x magnification was applied to a Haematoxylin colour deconvolved layer to isolate intensity peaks of haematoxylin staining corresponding to nuclei centroids. Following which all extraneous pixels were removed from further classification by applying a threshold of >25 to the LoG filtered layer. Dilation and hole filling functions were then used to further isolate relevant pixels into nuclear shapes, following which all resultant pixel clusters of >50μm^2 and <10μm^2 were

excluded. This function removed most large objects relating to tumour and stromal cell nuclei, and small areas / errant pixels not belonging to nuclear areas. Elongated nuclei were then excluded by applying an isoperimetric quotient, given by $\frac{4 \cdot \pi \cdot Area}{Perimeter^2}$, threshold of <0.7 to each pixel cluster, given that the isoperimetric quotient would be 1 for a perfect circle. The resultant classification contained a mix of lymphocytes and eosinophils, the latter of which was excluded by applying an eosin stain content threshold of <150 from a 5 x 5 maximum filtered, eosin colour deconvolved layer, encompassing >50% of the pixel cluster area.

2.4.6.2 Ki67 Positive Cell Detection

Cells positive for IHC stained Ki67 on both TMAs and full sections were detected using an algorithm adapted from a deep learning-based, brightfield nuclei detection algorithm developed by Visiopharm. The initial nuclei detection algorithm was developed by training a UNET CNN for 3 classes: Background, Nuclei, and Boundaries, in RGB colour space, and then removing the Background and Boundary classes, and objects <10µm^2 following tissue classification. The resultant Nuclei class was then separated into positive and negative cell classes by applying a to a DAB colour deconvolved layer threshold value of <150, encompassing >10% of the nuclear object area.

2.5 Statistical Analysis

Statistical analyses of survival metrics and comparative measures of statistical agreement were conducted in IBM SPSS version 25 (IBM, New York, USA). Processing of Visiopharm output files, data cleaning, generation of cut off values for all metrics and production of final figures was conducted in R Studio version 1.1.463 (RStudio, Boston, MA, USA).

3. Quantification of the Tumour Associated Stroma by Image Analysis in CRC

3.1 Introduction

In recent years, studies looking to elucidate methods for a precision medicine, patient-based approach to prognosis, have moved from studying the tumour in isolation to studying its symbiotic relationship with the Tumour Microenvironment (TME). The TME encompasses a variety of cell types and processes known to affect the development and spread of the tumour, such as tumour recruited neovasculature (Rmali et al., 2007) and host inflammatory cells (C. S. D. Roxburgh & McMillan, 2012). An element of the TME known to hold prognostic significance in multiple carcinomas is the tumour associated stroma, the supporting tissue of the tumour comprised of cancer associated fibroblasts (CAFs) and the extracellular matrix (ECM) they secrete (Freeman et al., 2013; Hasebe et al., 2001; Waghray et al., 2013).

Various studies have been conducted to determine of a method of assessment able to link the tumour associated stroma to prognosis. Early studies demonstrated a prognostic link between qualitative assessment of the proportion of stroma at the invasive margin and morphologic characterisation of the maturity of the stromal component (Halvorsen & Seimt, 1989; Ueno et al., 2004; Ueno, Jones, et al., 2002b). Recently however, the proportion of the stromal component to the tumour, the tumour-stroma ratio (TSR) or tumour-stroma percentage (TSP), has been repeatedly validated as an independent prognostic factor in CRC (Park et al., 2014b) and forms the basis of a prospective international reproducibility study for inclusion in routine pathology alongside the TNM system (M. Smit et al., 2019).

Assessment of the TSP on H&E sections is relatively uncomplicated; an FOV at 10x magnification representing the most stromally dense area with tumour cells at all four boundaries is graded in 5% (10% in some studies) increments, with a statistically determined cut off for high vs low patients at ~50%. Interobserver agreement in reporting high vs low stroma patients shows variable consistency, with Cohen's κ scores ranging from 0.5-0.97/1 for CRC, however, for clinical implementation, consistency in reporting is vital and thus further steps must be taken to ensure this.

Developments in machine learning as applied to histopathological image analysis, specifically the advent of CNNs for tissue segmentation, provide a logical path to addressing the issue of interobserver reporting variability, and indeed, efforts have been made to conduct the TSP assessment digitally (Geessink et al., 2019; Martin et al., 2020; Zhao et al., 2020). However, these studies utilise relatively low patient numbers and do not demonstrate translatability of the methodologies to geographical and process independent cohorts.

The aim of the present study, therefore, is to assess various image processing-based approaches to digitally conduct TSP assessment, namely traditional machine learning algorithms and deep learning CNN based methods, to assess the translatability of these methods to new cohorts, and to compare these methods to histopathological scoring, primarily in terms of ability to stratify patients by survival outcomes, and via statistical measures of inter-observer agreement.

3.2 Materials and Methods

3.1.1 Histopathological Assessment

Histopathological assessment of the TSP was carried out on H&E slides / WSI as described in 2.3.2 for each patient cohort. Briefly, an intratumoural area on the slide showing the deepest point of tumour invasion was assessed for percent of the FOV covered by desmoplastic stroma and scored in 5% increments with a cut – off of </> 50% stroma for high vs low TSP. The Norway patient cohort was assessed using the same methodology but only the TSP status (high vs low) was entered into the database. Additionally, the Norway cohort was scored by the same observer conducting the image analysis, all other sections were scored by an independent observer and obtained from existing databases.

3.2.2 Statistical Analysis

Patient cohort clinicopathological characteristics and test of association with CSS / DFS were determined in IBM SPSS version 25 (IBM, New York, USA). P values of <0.05 were considered to demonstrate a statistically significant association between clinicopathological features and clinical outcomes.

All other statistical analyses for this study were conducted in RStudio (RStudio, Boston, MA). Optimal cut off values for TSP scores were determined by using CSS / DFS as the endpoint with the maxstat and survminer packages. Receiver operator characteristic curves and associated AUC were plotted and calculated using the pROC package. Survival analysis was conducted and associated Kaplan – Meier survival curves produced using the survival, survminer, and survMisc packages, with the log rank statistic used to assess association between TSP stratification and CSS / DFS. Hazard ratios and associated 95% confidence intervals for survival analysis were calculated using the univariate Cox proportional hazard model.

3.3 Results

3.3.1 Automated Assessment of the Tumour Stroma Percentage by Machine Learning – Based Image Analysis

The initial approach to digitally quantify the TSP was conducted via a decision forest-based image segmentation algorithm utilising operator determined feature vectors for training and segmentation. The algorithm was initially trained in RGB colour space and on colour deconvolved haematoxylin and eosin layers, however resultant poor segmentation performance necessitated the generation of feature vectors that provided additional quantitative distinction between tumour and stroma to aid decision forest class weighting. A variety of features were generated and visually assessed for suitability in terms of segmentation, of which three were determined to provide the best quality of segmentation using this algorithm. The feature vectors (denoted F1, F2, F3, Figure 3.1), were primarily built on Haematoxylin and Eosin colour deconvolutions, given that in adenocarcinoma these can be considered the primary stains of the tumour and stroma respectively, with F1 designed to accentuate and smooth Haematoxylin dense regions such as nuclei in the tumour compartment, F2 designed to darken and smooth clusters of Haematoxylin dense regions but negate isolated nuclei, and F3 designed to accentuate nuclei at the periphery of tumour nests to provide a distinct boundary between the tumour and stroma. Representative images of the three feature vectors are shown in Figure 3.2.

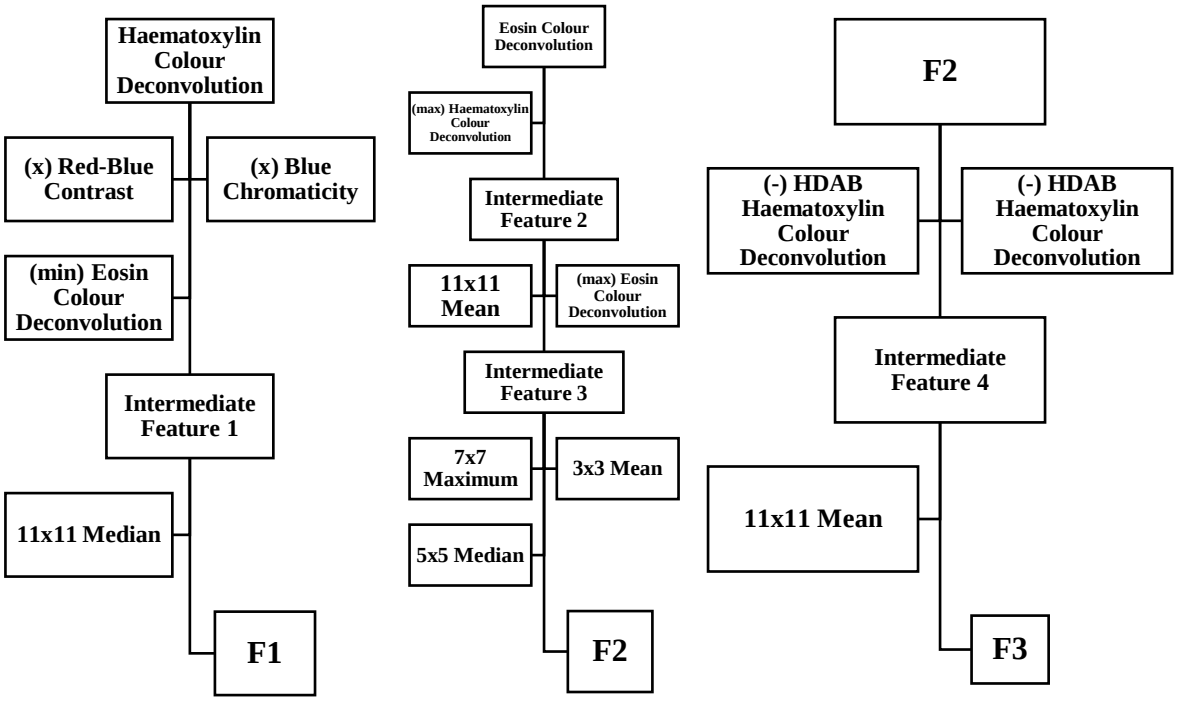


Figure 3.1 Features Utilised in Machine Learning TSP Algorithm. *Schematic demonstrating components and transformations of the three feature vectors generated for the machine learning TSP algorithm. Transformation steps are shown as: (transformation), transforming feature vector. Filtering steps are shown as: kernel size x kernel size, filter type.*

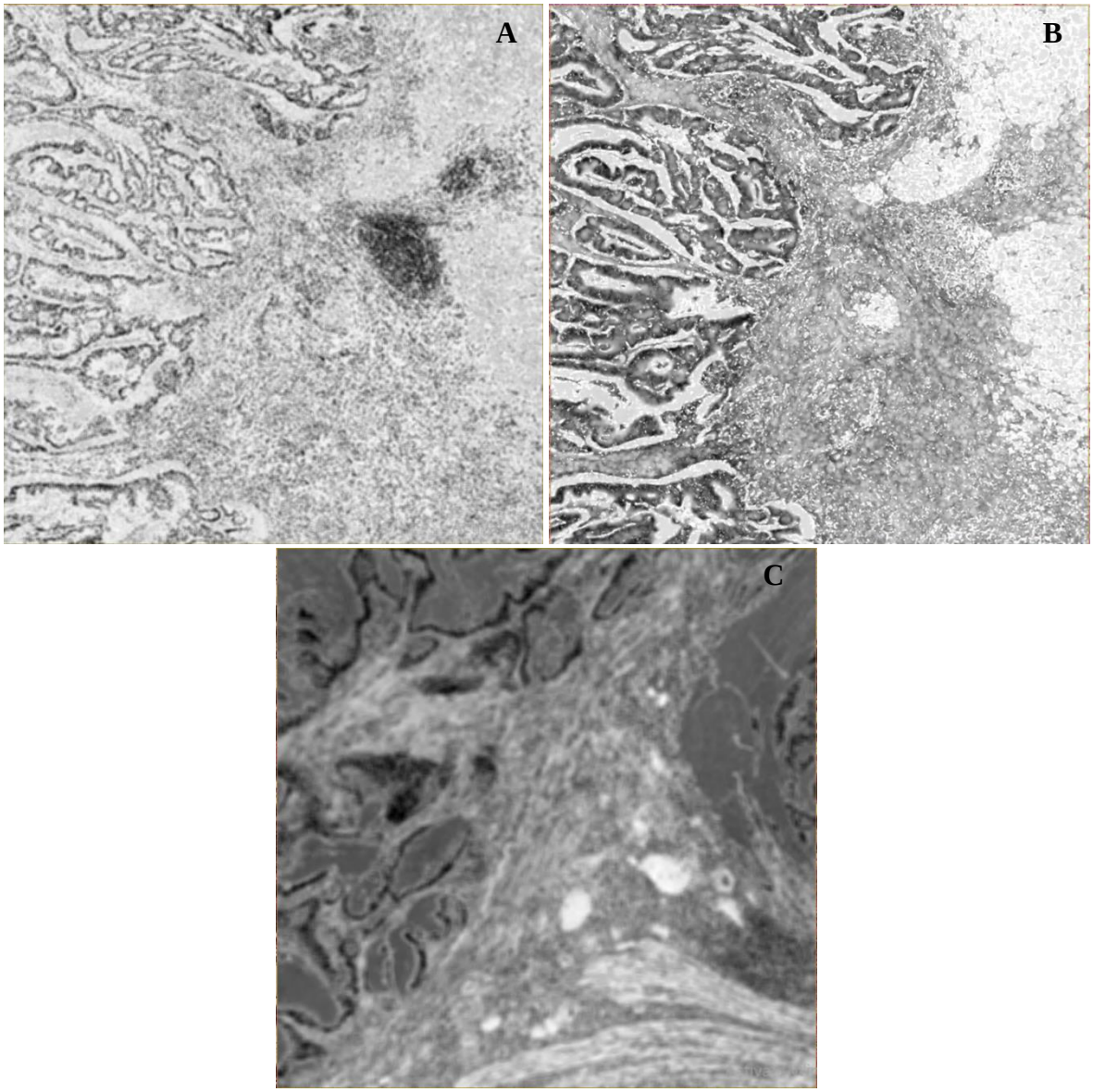


Figure 3.2 Representative Images of Feature Vectors Utilised in Machine Learning TSP Algorithm. *Representative grayscale images of the three feature vectors demonstrating the intended utility within the ML-based segmentation task. (A) F1, (B) F2, (C) F3.*

The algorithm was trained on training dataset 1 (detailed in 2.4.2.1) which comprised sixteen annotated WSI from the Norway cohort, 5% of the total cohort. A series of increasingly complex training parameters were tested on the three feature vectors detailed in Figure 3.1 to determine the optimum trade-off between manageable training times and quality of classification, the final parameters are detailed in 2.4.4.1.

During initial quality assessment of classification, two notable instances of algorithm confusion arose. Firstly, areas of high stain intensity and immune infiltrate in the stroma were misclassified as tumour, likely due to the use of haematoxylin colour deconvolution as the foundation of the feature vectors used to segment the tumour and stroma. To address this, the classified tissue was temporarily segmented into 100 μm^2 super pixels, F3 grayscale

values were normalized against the darkest 10% of the image, and super pixels with a mean intensity of lower than 500 of the normalized grayscale vectors were reclassified as stroma. Secondly, lymph node tissue particularly germinal centres, were consistently misclassified as tumour but proved more complicated to correct through post-classification processing. Therefore, prior to classification, lymph nodes in the subserosa were excluded (detailed in 2.4.3.2) from the analysis area as they were not to be utilized in any subsequent analysis.

In addition to automatic quantification of the TSP, the validated algorithm with the best clinical performance on all cohorts in this study was used to generate regions of interest for subsequent lymphocyte density analysis. Given that desmoplastic stroma can be present at a wide range of distances to the tumour on resection tissue, lymphocytes detected in these areas could not be guaranteed to be associated with an immunological response to the tumour. Nearchou and colleagues demonstrated that high densities of $CD3^+$ cells up to 100µm from tumour buds and the invasive margin in stage II CRC conferred a statistically significant improvement in survival (I. P. Nearchou et al., 2021; I. P. Nearchou et al., 2019). In addition, histopathological assessment was conducted on intratumoural FOV, therefore, to standardize subsequent analysis, only pixels classified as stroma and lymphoid tissue within 100µm of the tumour was quantified and converted to ROI for lymphocyte density analysis.

Finally, given that the algorithm conducts classification on a pixel-wise basis, the time required for conversions to ROI for subsequent analysis was augmented by non-contiguous areas with large numbers of pixels of multiple classes. Therefore, small pixel clusters ($<1000\mu m^2$) were converted to an intermediate class, then to the surrounding class with the largest neighbouring boundary interface length. The product of this operation was more homogenous classified areas but at a loss of some fine classification detail, notably intratumoural and stromal lymphoid areas where only a small number of lymphoid cells were present. However, given that intratumoural lymphocytes were not included in any subsequent analysis and stromal lymphocytes were to be quantified from the ROI generated by the validated TSP algorithm, this trade-off was considered acceptable and appropriate.

Image analysis for this study was conducted on WSI of the same H&E used to perform histopathological assessment. All WSI used for image analysis in this study were initially processed to remove all background artefacts and slide glass white space to produce a single continuous area for analysis around all viable tissue on the slide, detailed in 2.4.3.1. No colour space transformations or manual annotations were performed on the WSI prior to analysis. The TSP for each patient was calculated using the equation:

$\left(\frac{Area\ Stroma\ (\mu m^2) + Area\ Lymphoid(\mu m^2)}{Area\ Tumour(\mu m^2) + Area\ Stroma(\mu m^2) + Area\ Lymphoid(\mu m^2)}\right) \cdot 100$. The final image processing workflow for this algorithm is detailed in Figure 3.3 with representative images of the classifier performance demonstrated in Figure 3.4.

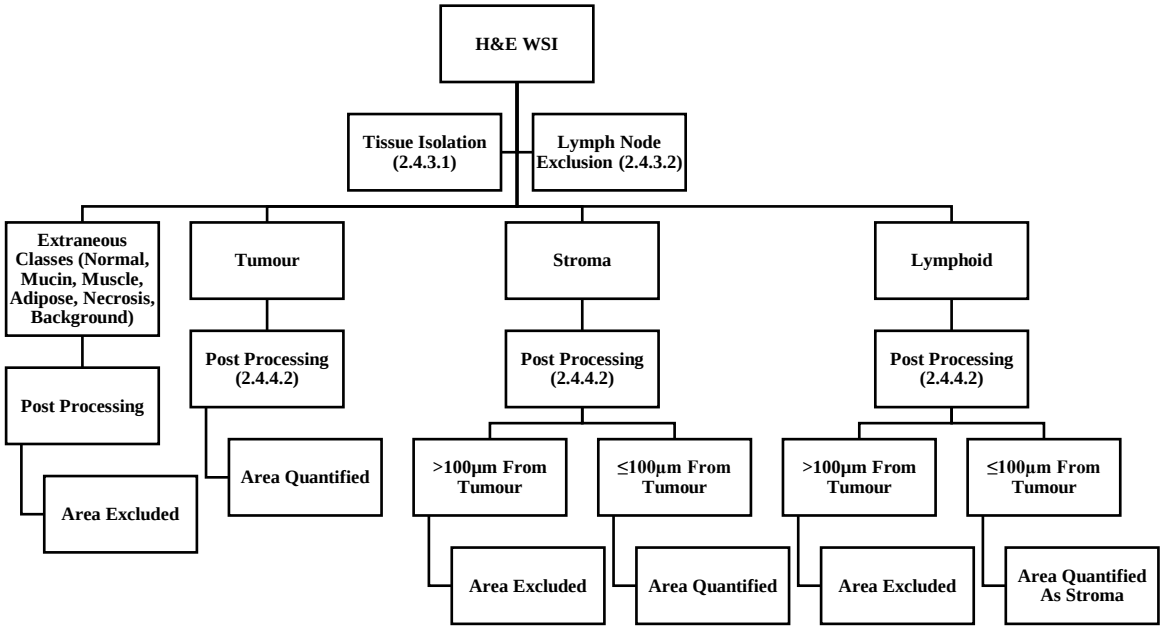


Figure 3.3 Machine Learning TSP Algorithm Workflow. *Workflow of ML TSP algorithm beginning with H&E WSI. Methods sections detailing operations are given in brackets.*

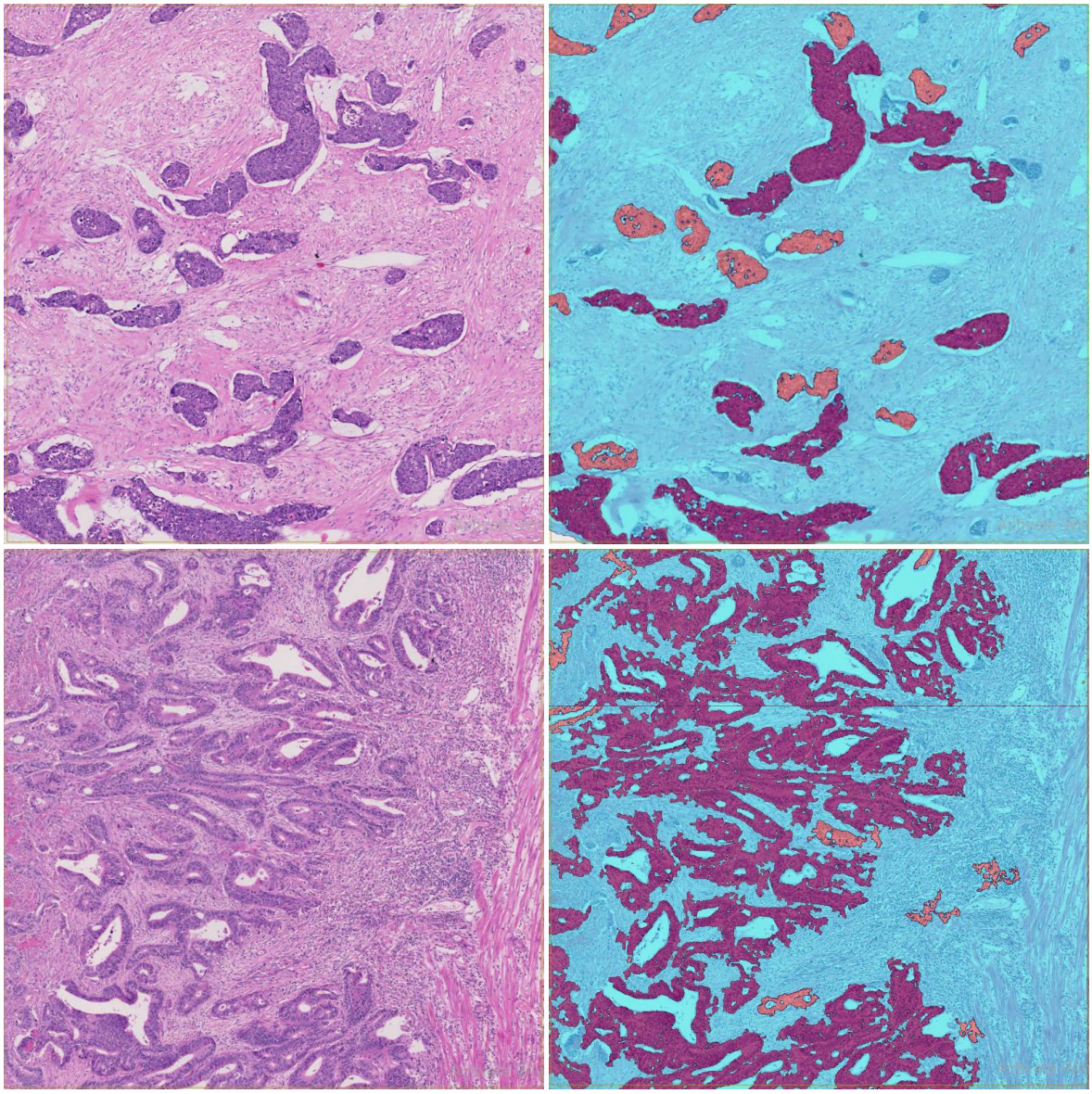


Figure 3.4 Representative Images of ML TSP Classifier Performance. *Images detailing the ML TSP classifier performance on sections from the Norway cohort. Tumour overlay is shown in red / orange and stroma in blue.*

Training and initial validation of this algorithm was conducted using a cohort of patients from Norway containing a total of 299 samples from Stage II – III CRC patients undergoing potentially curative resection in Kristiansand, Arendal, Flekkefjord, and Oslo Hospitals between 2005 – 2015. Patients had predominantly right sided (59.9%), T-Stage III (91.3%), node negative (63.2%) disease split evenly between sexes, with 52.5% being female. No patient included in this cohort received neoadjuvant chemotherapy or neoadjuvant radiotherapy and the majority (79.3%) did not receive adjuvant chemotherapy. Of the remaining patients, 7% received 5-Fluorouracil (5 – FU) only treatment and 13.7% received 5-FU in combination with Oxaliplatin, however, the administration of adjuvant chemotherapy did not have a significant effect on cancer specific survival (*P = 0.718).* No exclusion criteria were applied to this cohort as received; however, H&E tissue was missing for a single patient. The primary clinical outcome was CSS defined as the time from the date of surgery to death from primary disease. Clinicopathological features significantly associated with CSS were T-Stage (*P = 0.014)* and N-Stage (*P = 0.05)* (Table 3.1). The sixteen sections from this cohort that comprised training dataset 1, which the algorithm was trained on, were included in the final analysis due to the low number of CSS events in the cohort.

Clinicopathological Characteristic	N (%)	Clinical Outcome Significance
Sex		
Female	157 (52.5)	*0.556*
Male	142 (47.5)	
T Stage		
I	2 (0.7)	
II	10 (3.3)	*0.014*
III	273 (91.3)	
IV	14 (4.7)	
N Stage		
0	189 (63.2)	
1	83 (27.8)	*0.05*
2	27 (9.0)	
Tumour Site		
Right	179 (59.9)	
Left	31 (10.4)	*0.3*
Sigmoid	89 (29.8)	
Differentiation		
Well	3 (1.0)	
Moderate	225 (75.3)	*0.664*
Poor	57 (19.1)	
Adjuvant Chemotherapy		
None	237 (79.3)	
5-Fluorouracil only	21 (7.0)	*0.718*
5-FU + Oxaliplatin	41 (13.7)	

Table 3.1 Clinicopathological Characteristics of Norway Cohort. *Number (and %) of patients with clinicopathological features in Norway Cohort and association with Cancer Specific Survival.*

To assess the capability of the traditional machine learning algorithm to automatically score clinical H&E sections for the TSP, the machine learning tissue segmentation algorithm described above was applied to the Norway patient cohort H&E WSI. The TSP scores generated by the algorithm underwent ROC curve analysis to assess the predictive performance compared to the binary high / low TSP manual histopathological

scores, which were determined by the same observer conducting the image analysis. The AUC of the ROC curve was 0.794 (Figure 3.5).

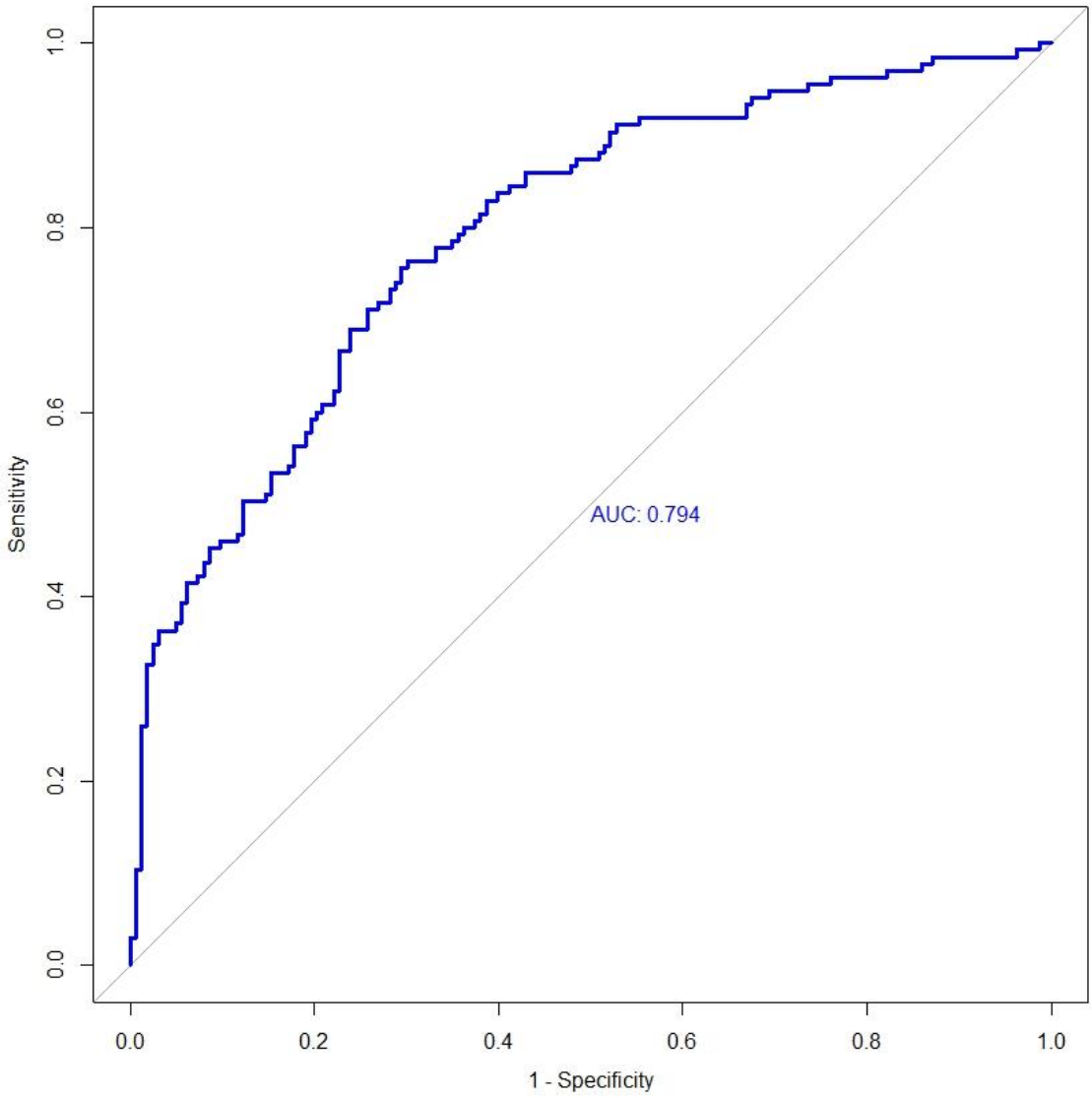


Figure 3.5 Receiver Operator Characteristic of ML generated TSP scores in Norway Cohort. *Machine learning determined TSP scores for 299 patients of Norway cohort compared to outcome of high vs low TSP status determined by histopathological assessment, AUC = 0.794.*

To determine an optimal cut off value to stratify patients into high vs low stroma based on CSS, the TSP scores generated by the algorithm were analysed using the survminer and maxstat packages, and an optimal cut point for CSS of 40.46% was determined (Figure 3.6).

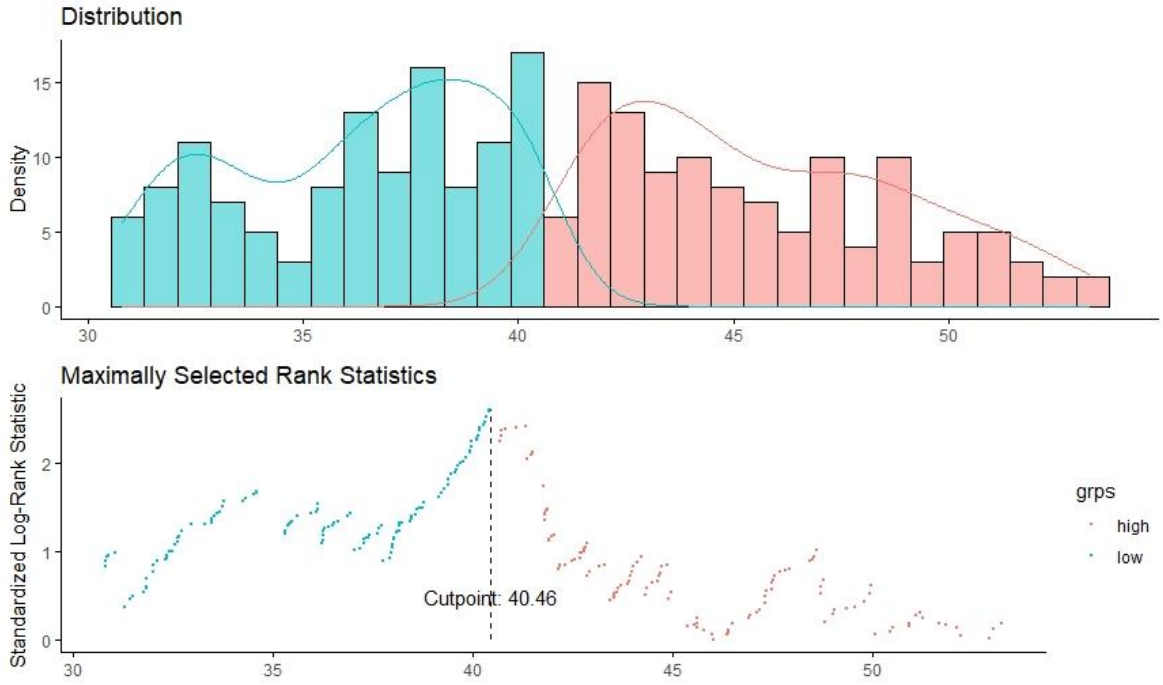


Figure 3.6 Determining optimal cut point for automated ML TSP in Norway Cohort. *Distributions and optimal cut point for high vs low stroma from ML TSP algorithm based on cancer specific survival from patients from Norway cohort. For this algorithm, the optimal cut point for CSS was determined to be 40.46%.*

Given that manual histopathological TSP assessment is conducted in 5% increments, for consistency the cut off value of 40.46% was rounded to the nearest 5% and set at 40%. This value was used to group patients into high vs low stroma and Kaplan – Meier survival analysis was performed to assess the algorithm's ability to stratify patients based on CSS. Consistent with previous work in other cohorts, manual histopathological TSP assessment (</> 50% Stromal Component) was significantly associated with CSS (HR 2.609, 95% CI 1.126 – 6.046, P = 0.02). Automated TSP assessment via machine learning (</> 40% Stromal Component) was also significantly associated with CSS (HR 2.259, 95% CI 0.9969 – 5.12, P = 0.045). Cohen's Kappa metric was utilised to assess the level of statistical agreement, showing moderate concordance between the two scores, $\kappa = 0.443$ (Figure 3.7).

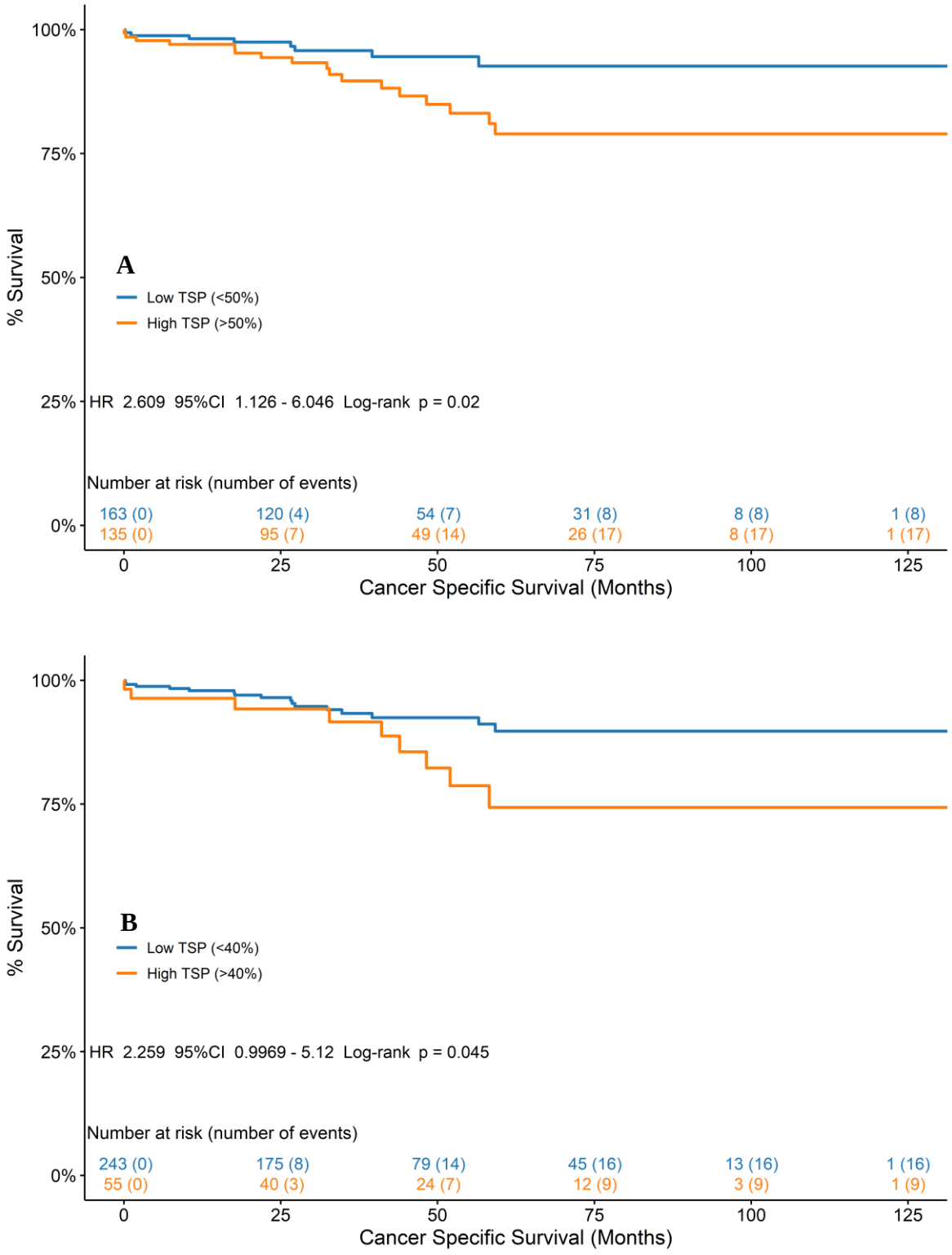


Figure 3.7 Relationship between TSP and CSS in Norway Cohort. *Association of cancer specific survival and TSP status determined via histopathological assessment (A) and machine learning based image analysis (B) in the Norway patient cohort. Cohen's Kappa agreement, $\kappa = 0.443$.*

3.3.2 Translation of Machine Learning – Based Image Analysis TSP Algorithm to a Geographically Independent Patient Cohort

Imperative to the clinical utility of image analysis is the ability of algorithms to perform on previously unseen slides and adapt to stochastic differences inherent to WSI from geographical and process independent cohorts. Therefore, to determine the ML algorithm's efficacy of assessment on a novel patient cohort, the same validation approach used in the Norway patient cohort was utilized in the Glasgow Development and Glasgow Validation patient cohorts.

The Glasgow Development Cohort comprised a total of 272 Stage I – III CRC patients undergoing potentially curative resection at the Glasgow Royal Infirmary Hospital between 1997 – 2007. The majority (68%) of patients presented with primary colon cancer split 39% and 29% between right sided and left sided tumours, respectively. There was a slightly larger proportion of males to females, 53.3% to 46.7% respectively, and a notably larger proportion over the age of 65 (63.7%). Primary disease was predominantly T-Stage III (61.0%) and node negative (56.4%), and the majority of patients (70.7%) did not receive adjuvant chemotherapy. Patients were excluded from survival analysis if they received neoadjuvant intervention or died within 30 days of surgery, leaving a total of 259 patients available for survival analysis. The primary clinical outcome was CSS defined as the time from the date of surgery to death from primary disease. Clinicopathological features associated with CSS were T-Stage ($P = 0.004$), and N-Stage ($P < 0.001$) (Table 3.2).

Clinicopathological Characteristic	N (%)	Clinical Outcome Significance
Age		
<65	94 (36.3)	*0.199*
>65	165 (63.7)	
Sex		
Female	121 (46.7)	*0.581*
Male	138 (53.3)	
T Stage		
I	8 (3.1)	
II	17 (6.6)	*0.004*
III	158 (61.0)	
IV	76 (29.3)	
N Stage		
0	146 (56.4)	
1	85 (32.8)	*<0.001*
2	28 (10.8)	
Tumour Site		
Right	101 (39.0)	
Left	75 (29.0)	*0.904*
Rectal	83 (32.0)	
Differentiation		
Well	10 (3.6)	
Moderate	235 (86.4)	*0.092*
Poor	26 (9.6)	
Adjuvant Chemotherapy		
Yes	76 (29.3)	*0.721*
No	183 (70.7)	

Table 3.2 Clinicopathological Characteristics of Glasgow Development Cohort. *Number (and %) of patients with clinicopathological features in Glasgow Development Cohort and association with Cancer Specific Survival.*

The Glasgow Validation Cohort comprised a total of 758 Stage I – IV CRC patients undergoing potentially curative resection at the Glasgow Western Infirmary and Stobhill Hospitals between 2000 – 2007 identified from a retrospective database. Patients were predominantly over the age of 65 (71.4%), presenting primarily with right sided (43.3%), T-Stage III (52.9%), node negative (65.3%) disease split evenly between sexes (50.5% male). Patients were excluded from survival analysis if they were administered neoadjuvant therapy, died within 30 days of surgery, or presented with Stage – IV disease. This left 654 patients available for survival analysis, of which, 23 had missing or inappropriate H&E tissue for image analysis, meaning 631 patients were included in the final study. The primary clinical outcome was CSS defined as the time from the date of surgery to death from primary disease. Clinicopathological features associated with CSS were T-Stage (*P <0.001)*, and N-Stage (*P <0.001)* (Table 3.3).

Clinicopathological Characteristic	N (%)	Clinical Outcome Significance
Age		
<65	187 (28.6)	*0.990*
>65	467 (71.4)	
Sex		
Female	324 (49.5)	*0.475*
Male	330 (50.5)	
T Stage		
I	32 (4.9)	
II	92 (14.1)	*<0.001*
III	346 (52.9)	
IV	184 (28.1)	
N Stage		
0	425 (65.3)	
1	152 (23.3)	*<0.001*
2	74 (11.4)	
Tumour Site		
Right	325 (43.3)	
Left	263 (35.0)	*0.544*
Rectal	163 (21.7)	
Differentiation		
Well	28 (3.7)	
Moderate	650 (85.8)	*<0.001*
Poor	68 (9.0)	
Adjuvant Chemotherapy		
Yes	28 (48.3)	*0.434*
No	30 (51.7)	

Table 3.3 Clinicopathological Characteristics of Glasgow Validation Cohort. *Number (and %) of patients with clinicopathological features in Glasgow Validation Cohort and association with Cancer Specific Survival.*

Manual histopathological assessment of the TSP in the Glasgow Development Cohort was conducted by Jamie Park as part of the original study validating the TSP in this cohort (Park et al., 2014), and by Antonia Roseweir for the Glasgow Validation Cohort as part of a multiple cohort validation study (Roseweir et al., 2020). The predictive performance of the TSP values determined by the algorithm in the novel cohorts was assessed via ROC curve analysis, with the Glasgow Development Cohort producing an AUC of 0.590 and the Glasgow Validation Cohort 0.567 (Figure 3.8).

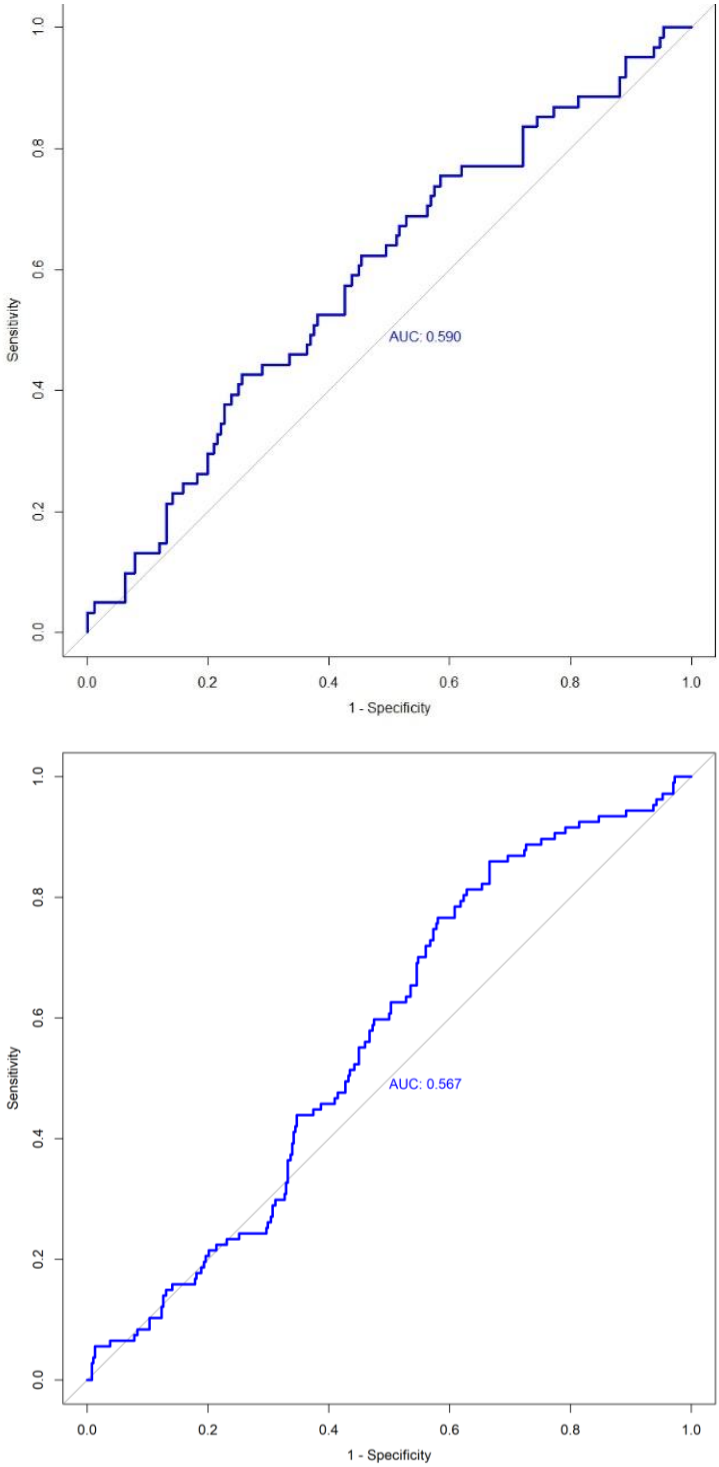


Figure 3.8 Receiver Operator Characteristic of ML generated TSP scores in Glasgow Development Cohort. *Machine learning determined TSP scores for 259 patients of Glasgow Development Cohort (top) and 631 patients of the Glasgow Validation Cohort (bottom) compared to outcome of high vs low TSP status determined by histopathological assessment, AUC = 0.590 & 0.567 respectively.*

To validate the cut off value of 40% stroma determined from the Norway cohort, prior to survival analysis the same maximally ranked statistics analysis was applied to the TSP scores generated in the novel cohorts, with a value of 40.67% stroma obtained for the Glasgow Development Cohort and 40.56% for the Glasgow Validation Cohort (Figure 3.9).

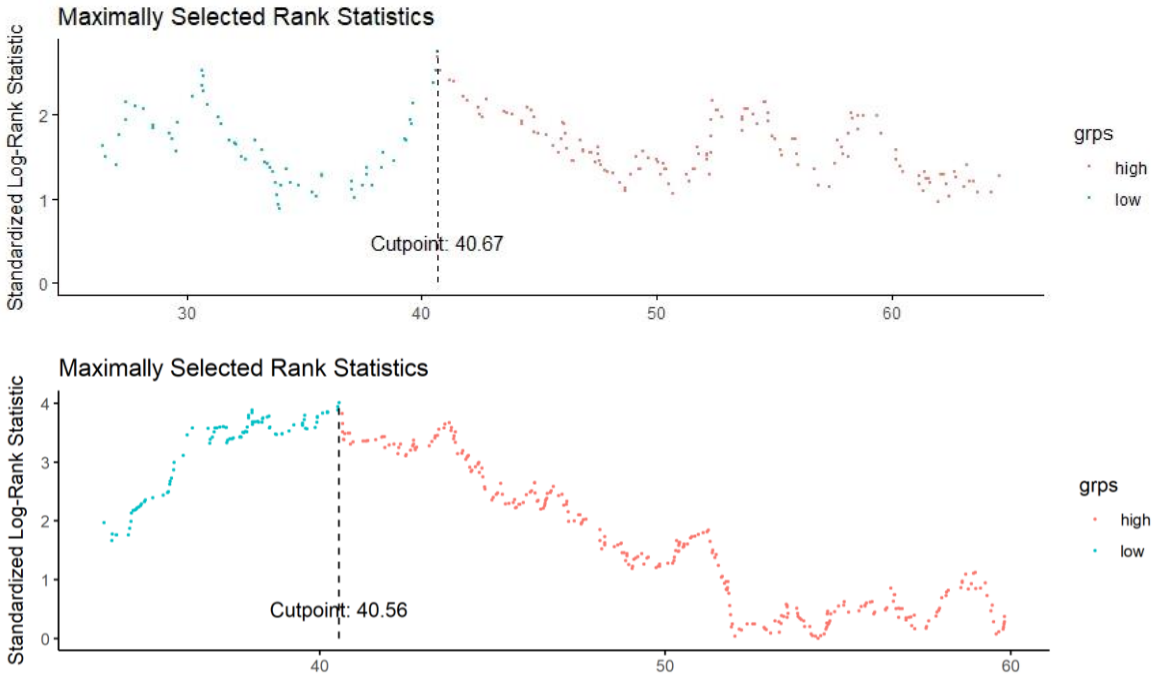


Figure 3.9 Validating optimal cut point for automated ML TSP in Glasgow Development and Validation Cohorts. *Distributions and optimal cut point for high vs low stroma from ML TSP algorithm based on cancer specific survival from patients from Glasgow Development (top) and Validation (bottom) cohorts. For this algorithm applied to the novel patient cohorts, the optimal cut off value was 40.67% and 40.56% respectively.*

Given the similarity in cut off values and in the interest of translatability, the cut off value of </>40% defined in the Norway cohort was used in the Glasgow Development and Validation cohorts to stratify patients into high vs low stroma. Kaplan-Meier survival analysis showed a very strong trend towards association with CSS in the Glasgow Development cohort but failed to reach statistical significance (HR 1.559 95% CI 0.9657 – 2.516 P = 0.067, Figure 3.10), and a moderate association with CSS in the Glasgow Validation cohort (HR 1.359 95% CI 0.8529 – 2.167 P = 0.2, Figure 3.11). Cohen's Kappa metric was utilised to assess the level of statistical agreement, showing slight concordance between the two scores in both cohorts, Glasgow Development $\kappa = 0.151$ and Glasgow Validation $\kappa = 0.11$ (Figure 3.10 & 3.11, respectively).

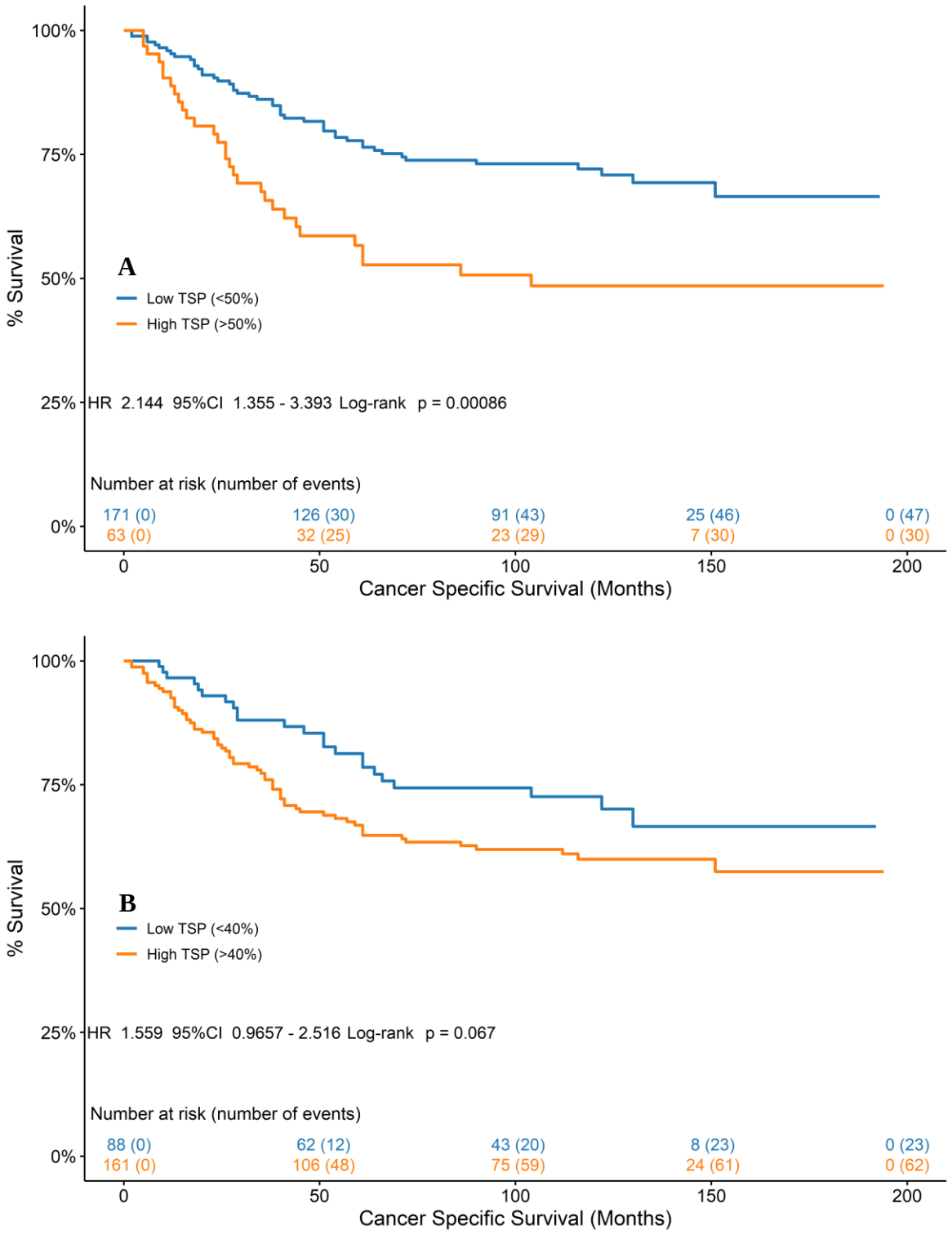


Figure 3.10 Relationship between CSS and TSP in Glasgow Development cohort. *Association of CSS and TSP determined via histopathological assessment (A) and machine learning based image analysis (B) in the Glasgow Development patient cohort. Cohen's Kappa agreement, κ = 0.151.*

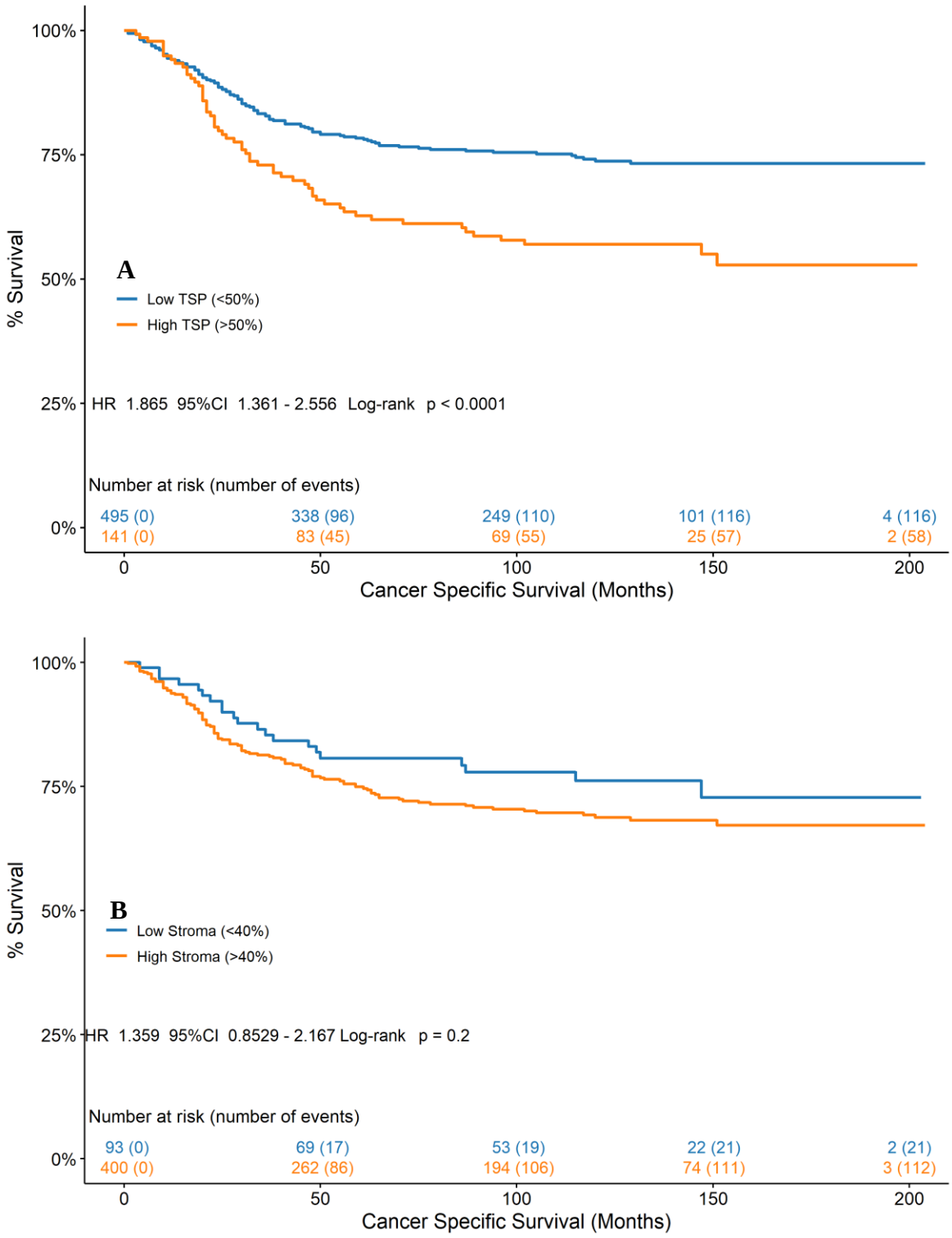


Figure 3.11 Relationship between CSS and TSP in Glasgow Validation cohort. *Association of CSS and TSP determined via histopathological assessment (A) and machine learning based image analysis (B) in the Glasgow Validation patient cohort. Cohen's Kappa agreement, $\kappa = 0.11$.*

3.3.3 Automated Assessment of The Tumour Stroma Percentage by Deep Learning – Based Image Analysis

The advent of CNNs for semantic image segmentation, and the inherent capability of these algorithms to adapt to new data, has starkly increased the translational capacity of image analysis in a pathology setting. This offers a much more promising foundation to develop an image processing algorithm that is able to accurately classify tissue across multiple geographical and process independent cohorts, therefore, a series of CNN based methodologies were investigated to determine the optimum conditions for automated TSP assessment. Given the ability of CNNs to handle much larger and more complex data sets, a training data set comprised of 7180 224 x 224-pixel image patches of the CRC-VAL-HE-7K dataset (Kather et al., 2019, 2.4.2.2) was constructed to provide a significantly greater number of data points for each tissue class. Additionally, the image patches from this dataset underwent Macenko's colour normalization (Macenko et al., 2009) thus reducing the possible effect of any histological stain variation present in the manually annotated data.

Given that all image analysis for this study was conducted in the Visiopharm software and for ease of use, the neural network architecture was selected from one of the three available in the deep learning module of the software; FCN-8s, DeepLabV3+, and U-NET. FCN-8s is a fully convolutional implementation of the VGG-16 net architecture that combines predictions from the final output with progressively earlier pooling function outputs to predict fine detail while retaining high level semantic information (Shelhamer et al., 2017; Simonyan & Zisserman, 2015). DeepLabV3+ is an extension of the DeepLabV3 model, that uses the DeepLabV3 as the encoder in an encoder-decoder structure. DeepLabV3+ employs the Xception model as the network backbone and adds a decoder module to improve the segmentation notably along object boundaries (Chen et al., 2018; Chollet, 2017). U-NET is a network architecture designed specifically for semantic segmentation of biomedical images, that utilises concatenation functions between the corresponding cropped images of the contracting and expansive paths to greatly improve boundary detection (Ronneberger et al., 2015c).

Each network architecture was trained on Training Data Set 2 (2.4.2.2) with weights generated from pre-training on the ImageNet dataset, utilising the same hyperparameters (512 x 512 image input, mini-batch size = 8, learning rate = 10^{-4}) with cross entropy + Intersection Over Union (IOU) loss monitored until a consistent low rate of <0.5 was achieved, meaning training ranged from 100 – 200 epochs dependent upon the network. The trained networks were then visually assessed for quality and speed of segmentation across

multiple sections; with U-NET consistently outperforming the other two networks it was chosen to be the model taken forward for validation of TSP assessment on the clinical samples.

Following quantification of the TSP, the validated algorithm was to be used for subsequent cellular analysis that required class overlays generated by the algorithm to be converted into ROI. The process of pixel-wise segmentation produces non-homogenous overlay areas across the tissue that can substantially augment the time required to run overlay to ROI conversions, therefore, post-processing steps were included in the segmentation algorithm to smooth the overlay and reduce the subsequent computational burden. To do this, the same operation used in 3.3.1 was utilised in this algorithm, whereby pan – class pixel clusters of < 1000μm^2 were converted into an intermediate class and then converted into the class with which they shared the largest neighbouring boundary interface length. The result of this was similar to the ML algorithm in that single lymphocytic cell detail was lost primarily to the stromal compartment, however, given that these areas were to be converted to ROI for subsequent lymphocyte cell density analysis and lymphocytes are part of the cellular component of the tumour microenvironment and associated stroma, this was considered acceptable. No further post-processing operations were performed but prior to the algorithm being applied to the sections, the same tissue isolation step used in the ML approach was applied (2.4.3.1), and under the same rationale as the ML algorithm, only stroma and lymphoid pixels within 100μm of the tumour was quantified as part of the TSP. The finalised DL TSP algorithm workflow is shown in Figure 3.12 and representative images of classifier performance are shown in Figure 3.13.

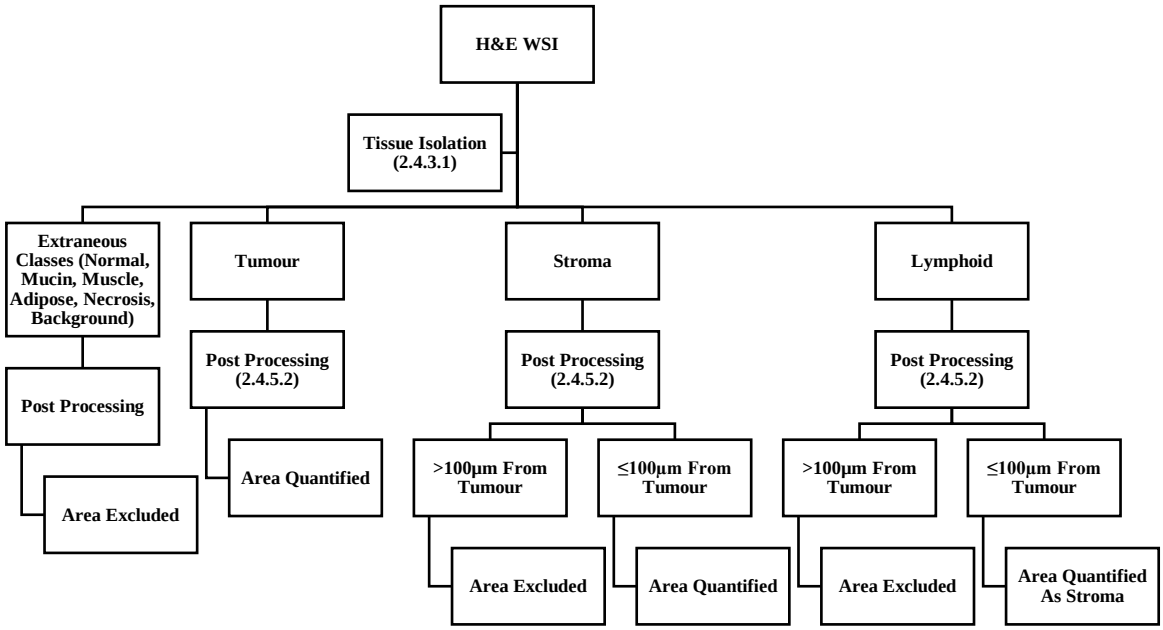


Figure 3.12 U-NET TSP Algorithm Workflow. *Workflow of DL U-NET TSP algorithm beginning with H&E WSI. Methods sections detailing operations are given in brackets.*

To validate the performance of the deep learning TSP algorithm, 10 sections were selected from the Norway, Glasgow Development and Glasgow Validation Cohorts and exhaustively annotated for the tumour and stroma. The TSP algorithm, including post processing, was then applied to the slides to determine the Intersection Over Union (IoU) score for both tissue compartments. The algorithm demonstrated excellent concordance with the tumour classification, with an average IoU of 0.911 and a maximum of 0.980, and good concordance with stroma classification, with an average IoU of 0.813 and a maximum of 0.927. Overall, tumour classification showed greater consistency of classification, IoU range 0.836-0.980, than stroma classification, IoU range 0.624-0.927, Table 3.4.

Slide / Tissue	1	2	3	4	5	6	7	8	9	10	Average
Tumour	*0.875*	***0.980***	*0.933*	*0.925*	*0.911*	*0.836*	*0.910*	*0.875*	*0.953*	*0.911*	***0.911***
Stroma	*0.886*	*0.758*	*0.861*	*0.790*	*0.624*	***0.927***	*0.816*	*0.819*	*0.831*	*0.817*	***0.813***

Table 3.4 Intersection over Union for Tumour / Stroma Classification by Deep Learning TSP algorithm. IoU values for Tumour and Stroma classification on 10 exhaustively annotated H&E WSI. Highest individual IoU for each tissue compartment in bold.

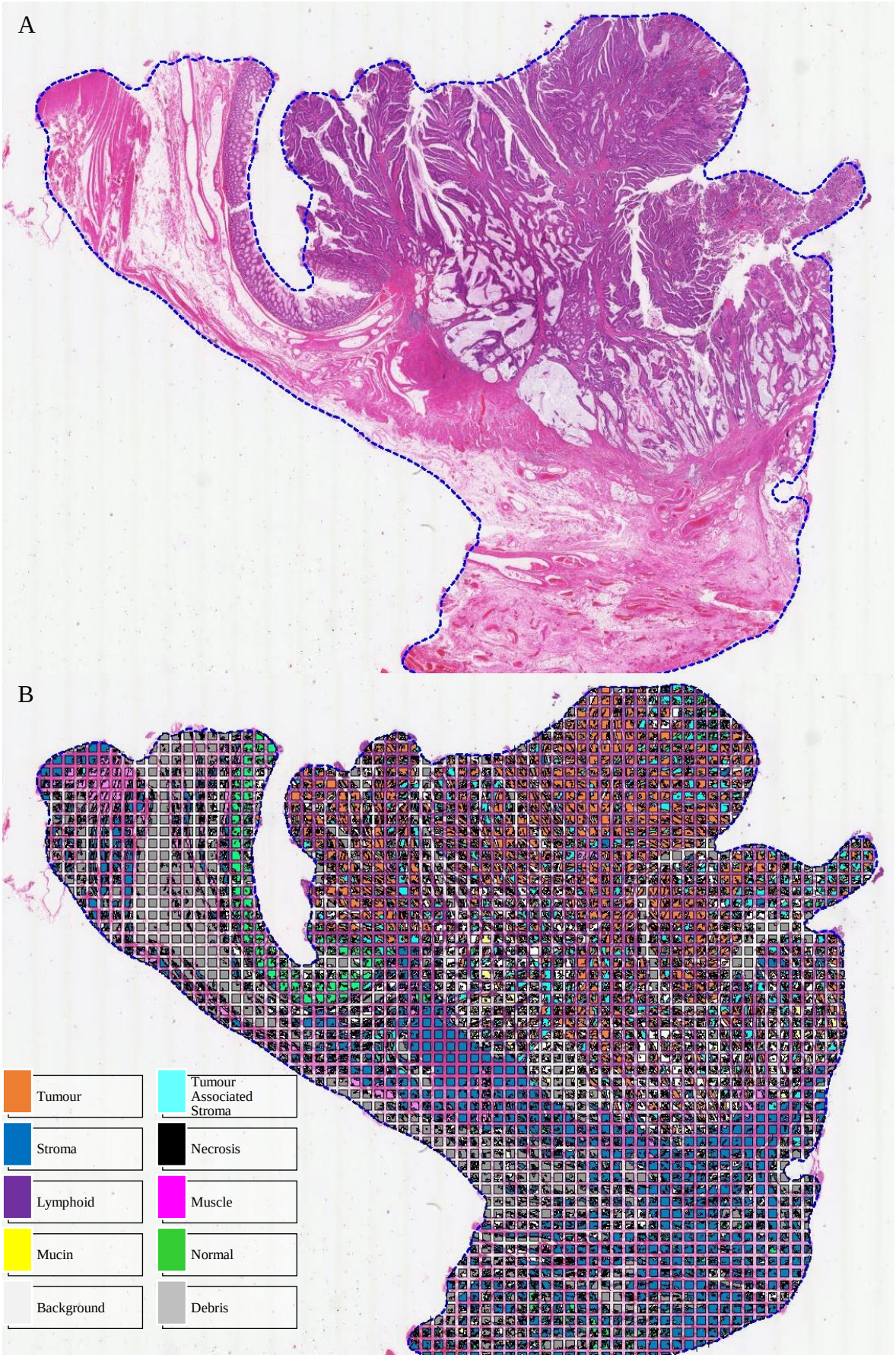


Figure 3.13 Deep Learning Classifier Performance on WSI. *Continued on next page.*

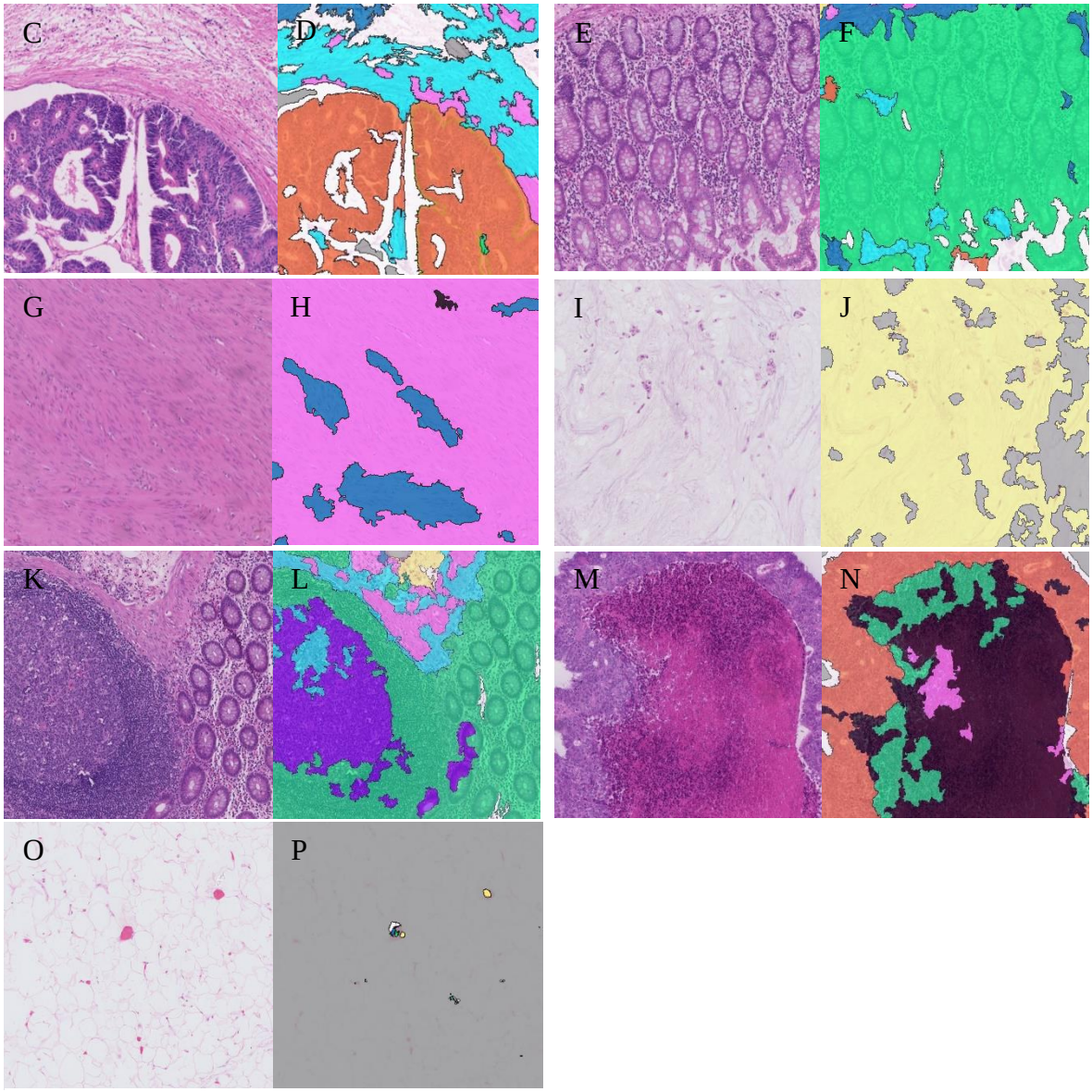


Figure 3.13 Deep Learning Classifier Performance on WSI. *Representative images of deep learning algorithm performance. WSI (A) and classification overlay (B) demonstrating whole section segmentation. CEGIKMO, detail images of algorithm performance on Tumour / Stroma (C), Benign Epithelium (E), Muscle (G), Mucin (I), Lymphoid Tissue (K), Necrosis (M), and Mucin (O). DFHJLNP, Algorithm classification overlay for respective images.*

Initial validation of the DL TSP algorithm was conducted as per the ML TSP algorithm in the same Norway Patient cohort. The TSP value for each patient was calculated from the areas of each relevant class determined by the algorithm using the equation: $\left(\frac{Area\ Stroma\ (\mu m^2) + Area\ Lymphoid(\mu m^2)}{Area\ Tumour(\mu m^2) + Area\ Stroma(\mu m^2) + Area\ Lymphoid(\mu m^2)}\right) \cdot 100$. The scores underwent ROC curve analysis to determine the predictive quality of the algorithm and produced and AUC of 0.834 for this cohort (Figure 3.14).

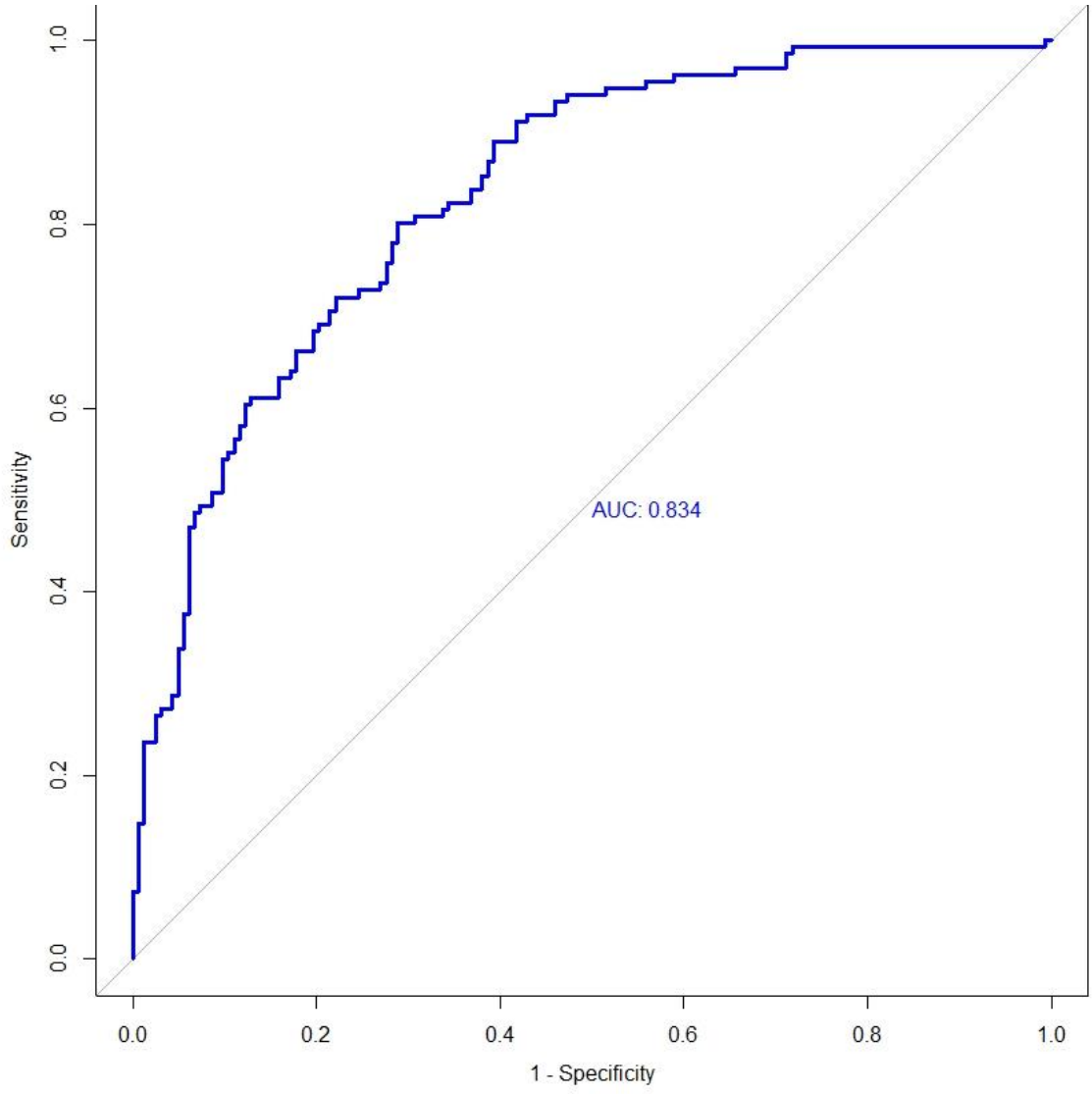


Figure 3.14 Receiver Operator Characteristic of DL generated TSP scores in Norway Cohort. *Deep learning determined TSP scores for 299 patients of Norway cohort compared to outcome of high vs low TSP status determined by histopathological assessment, AUC = 0.834.*

Algorithm generated TSP scores were analysed against CSS to determine an optimal cut off value to group patients into high vs low stroma. The maxstat and survminer packages were used to generate the cut off value, which was 41.35% for this algorithm in this cohort (Figure 3.15).

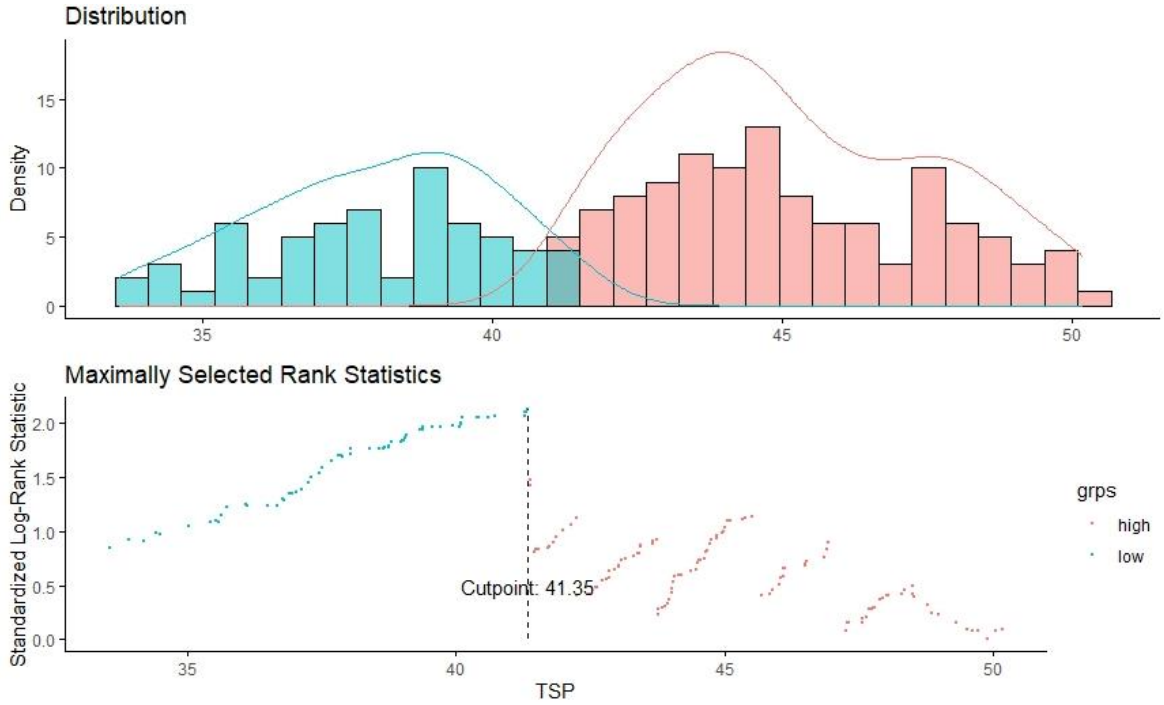


Figure 3.15 Determining optimal cut point for DL automated TSP in Norway Cohort. *Distribution of DL automated TSP scores and optimal cut point for high vs low stroma based on cancer specific survival in Norway patient cohort. For this algorithm in this cohort, the optimal cut off value was 41.35%.*

As per the ML algorithm, the cut off value determined for the DL algorithm in the Norway cohort was rounded to the nearest 5% to stratify patients into high vs low stroma for survival analysis, meaning a cut off value of 40% was used again for this study. Kaplan – Meier survival analysis was performed on the stratified DL generated TSP scores to determine association with CSS and compare survival stratification performance against manual histopathological scores. The DL generated TSP showed a statistically significant association with CSS in the Norway cohort (HR 3.178 95% CI 1.269 – 7.96, P = 0.0091) (Figure 3.16), outperforming the TSP stratification generated by the ML algorithm. Cohen's Kappa metric was utilised to assess the level of statistical agreement, showing improved moderate concordance between the two scores over the ML algorithm, $\kappa = 0.471$ (Figure 3.16).

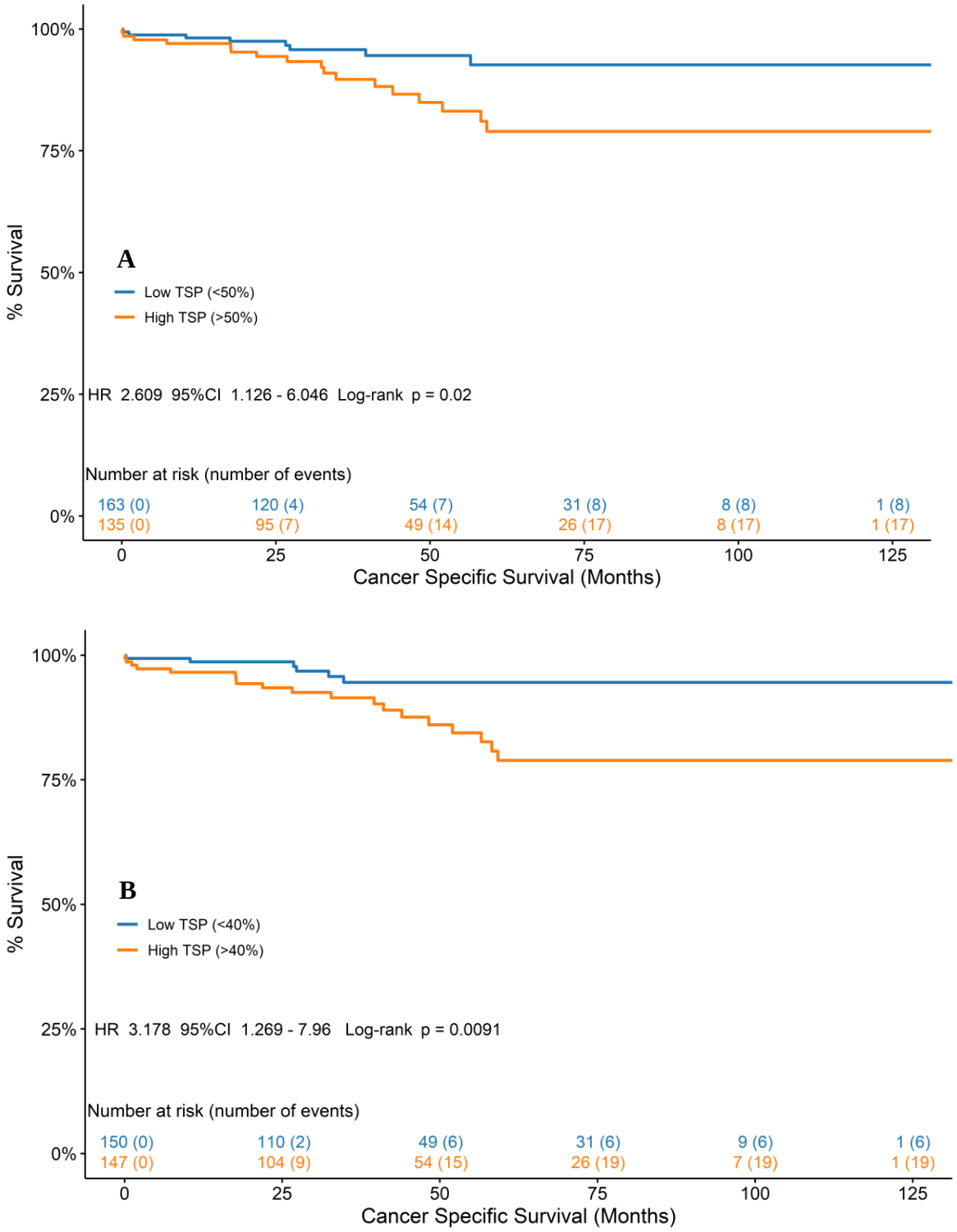


Figure 3.16 Relationship between TSP and CSS in Norway Cohort. *Association between CSS and TSP scores determined via histopathological assessment (A) and DL based image analysis (B) in Norway patient cohort. Cohen's Kappa agreement, κ = 0.471.*

3.3.4 Translation of Deep Learning – Based Image Analysis TSP Algorithm to Geographically Independent Patient Cohorts

To assess the capability of the DL TSP algorithm to accurately assess the TSP in geographical and process independent cohorts, the algorithm was applied to the same two validation cohorts as the ML algorithm, the Glasgow Development and Validation cohorts. The TSP scores generated for each patient underwent ROC analysis against the histopathological assessment, with the Glasgow Development cohort producing and AUC = 0.691 and the Glasgow Validation cohort AUC = 0.720 (Figure 3.17).

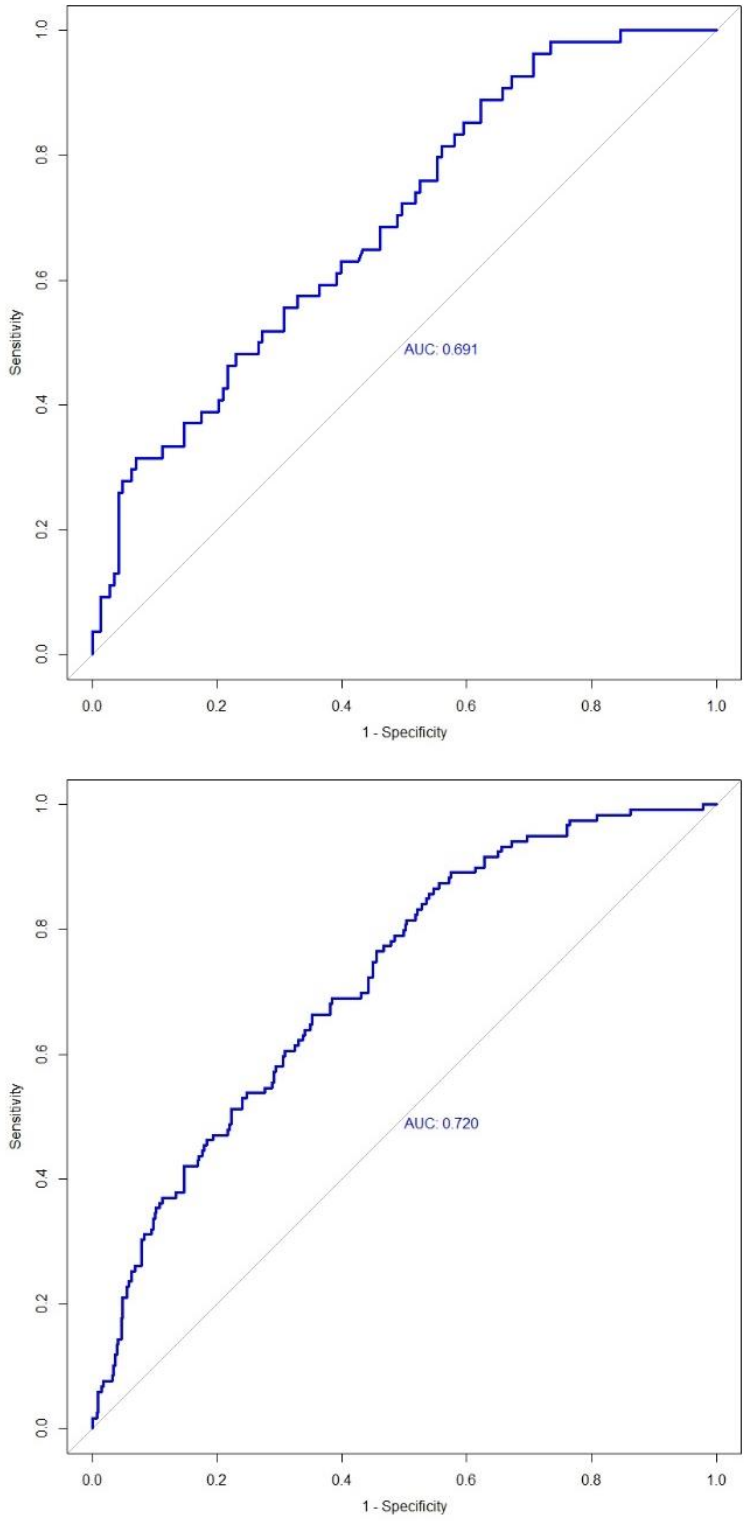


Figure 3.17 Receiver Operator Characteristics of DL Generated TSP in Geographically Independent Patient Cohorts. *ROC curves for DL generated TSP scores against histopathological TSP status in two validation CRC patient cohorts. (top) Glasgow Development Cohort, AUC = 0.691. (bottom) Glasgow Validation Cohort, AUC = 0.720.*

Validation of the DL TSP algorithm by Kaplan – Meier survival analysis showed a statistically significant association with CSS in the Glasgow Development Cohort (HR 2.265 95% CI 1.439 – 3.566, P = 0.00034, Figure 3.18) and Glasgow Validation Cohort (HR 1.559 95% CI 1.12 – 2.169, P = 0.0079, Figure 3.19). Cohen's Kappa metric was utilised to assess the level of statistical agreement, showing fair concordance between the two scores in the Glasgow Development cohort, $\kappa = 0.21$, and slight concordance in the Glasgow Validation cohort, $\kappa = 0.18$ (Figure 3.18 and 3.19, respectively). This demonstrates that the DL algorithm translates between cohorts for survival prediction but does not significantly correlate with pathologists assessment.

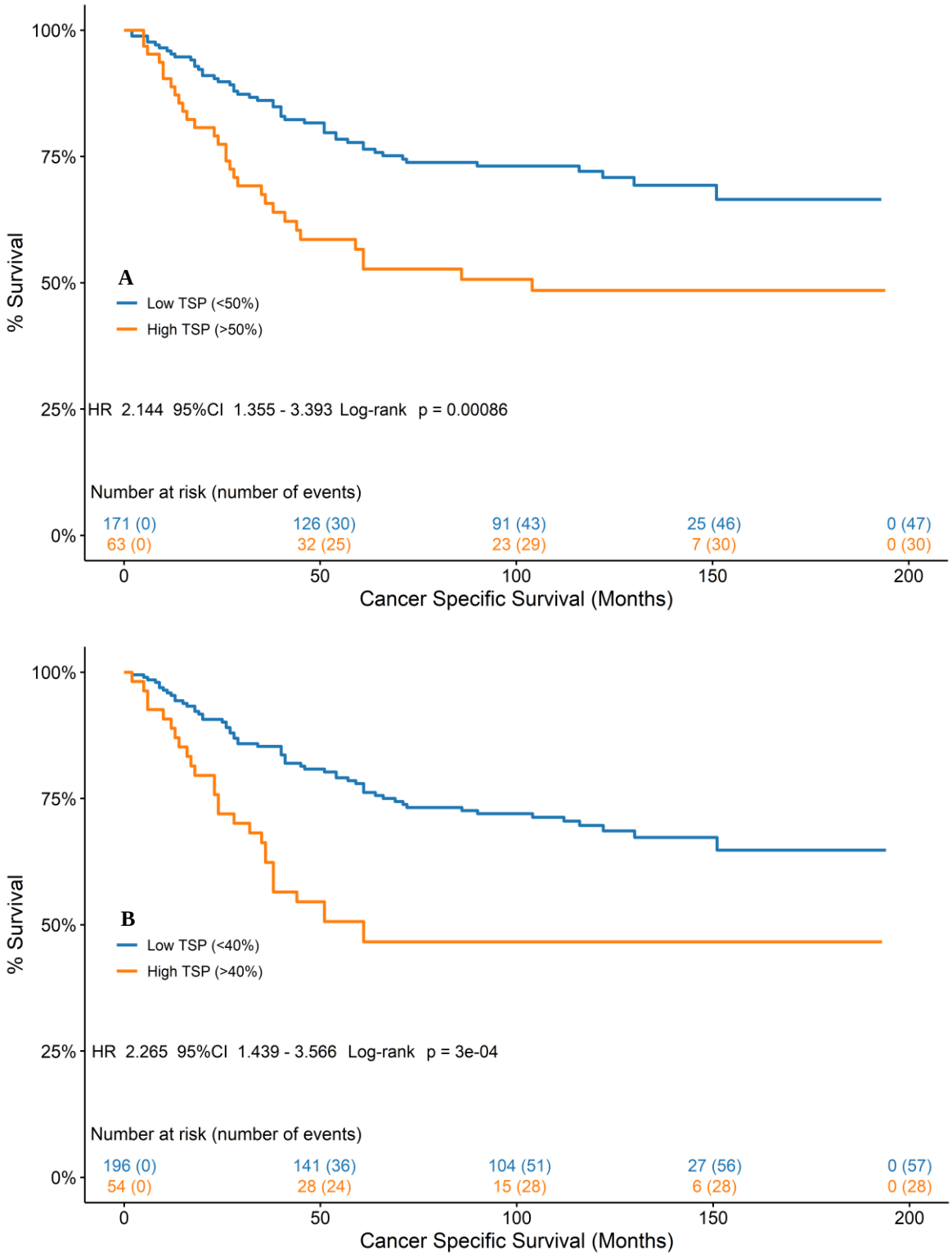


Figure 3.18 Relationship between CSS and TSP in Glasgow Development Cohort. *Association of CSS and TSP determined by (A) histopathological assessment (HR 2.144 95% CI 1.355 – 3.393, P = 0.00086) and (B) automated DL assessment (HR 2.265 95% CI 1.439 – 3.566, P = 0.00034) in Glasgow Development Cohort. Cohen's Kappa agreement, κ = 0.21.*

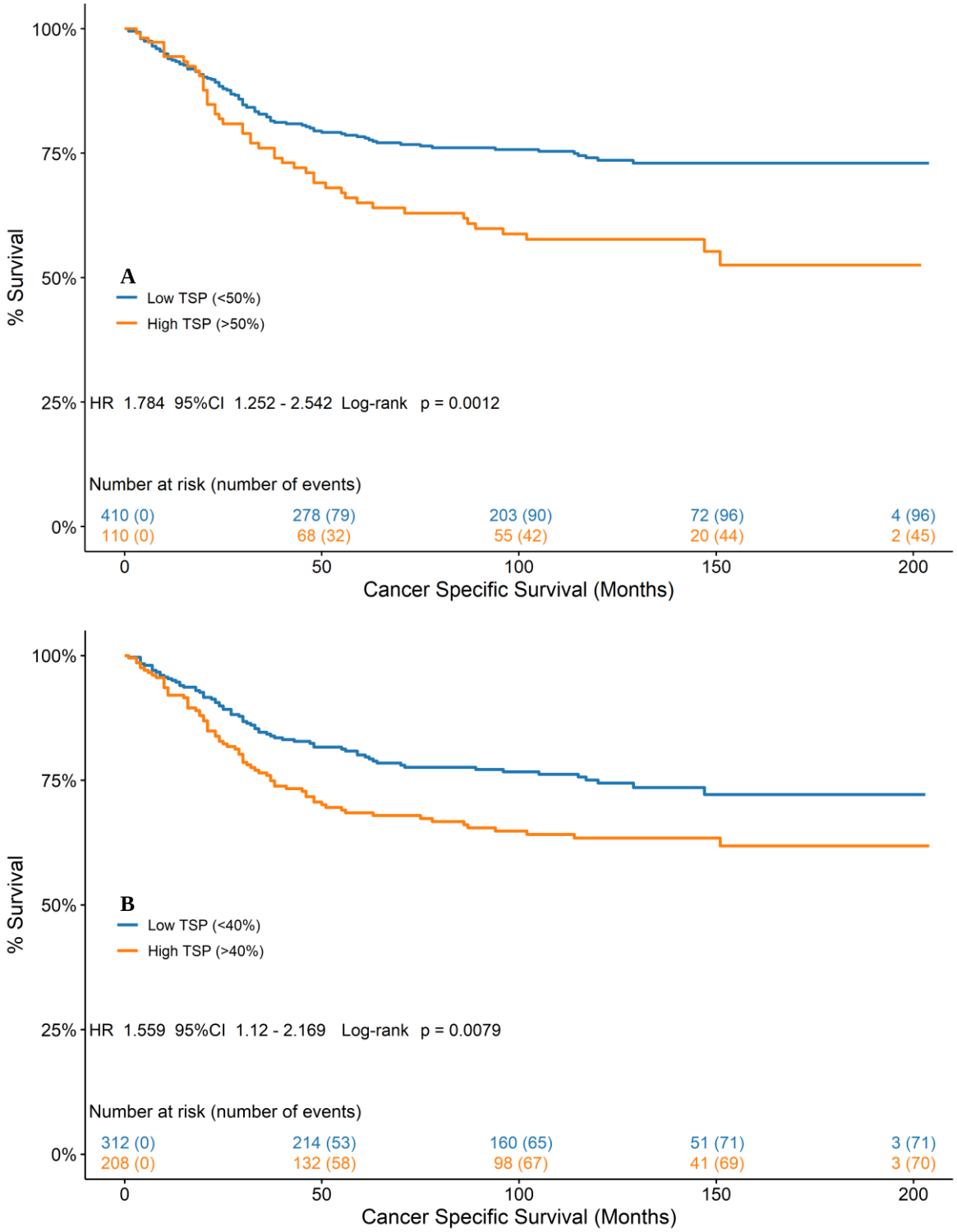


Figure 3.19 Relationship between CSS and TSP in Glasgow Validation Cohort. *Association of CSS and TSP determined by (A) histopathological assessment (HR 1.865 95% CI 1.361 – 2.556, P <0.0001) and (B) automated DL assessment (HR 1.559 95% CI 1.12 – 2.169, P = 0.0079) in Glasgow Validation Cohort***.** *Cohen's Kappa agreement, κ = 0.18.*

Bland Altman plots were constructed for each cohort to assess the relationship between the manual histopathological and algorithm generated TSP scores, and to assess any potential systemic bias in the scoring (Figure 3.20).

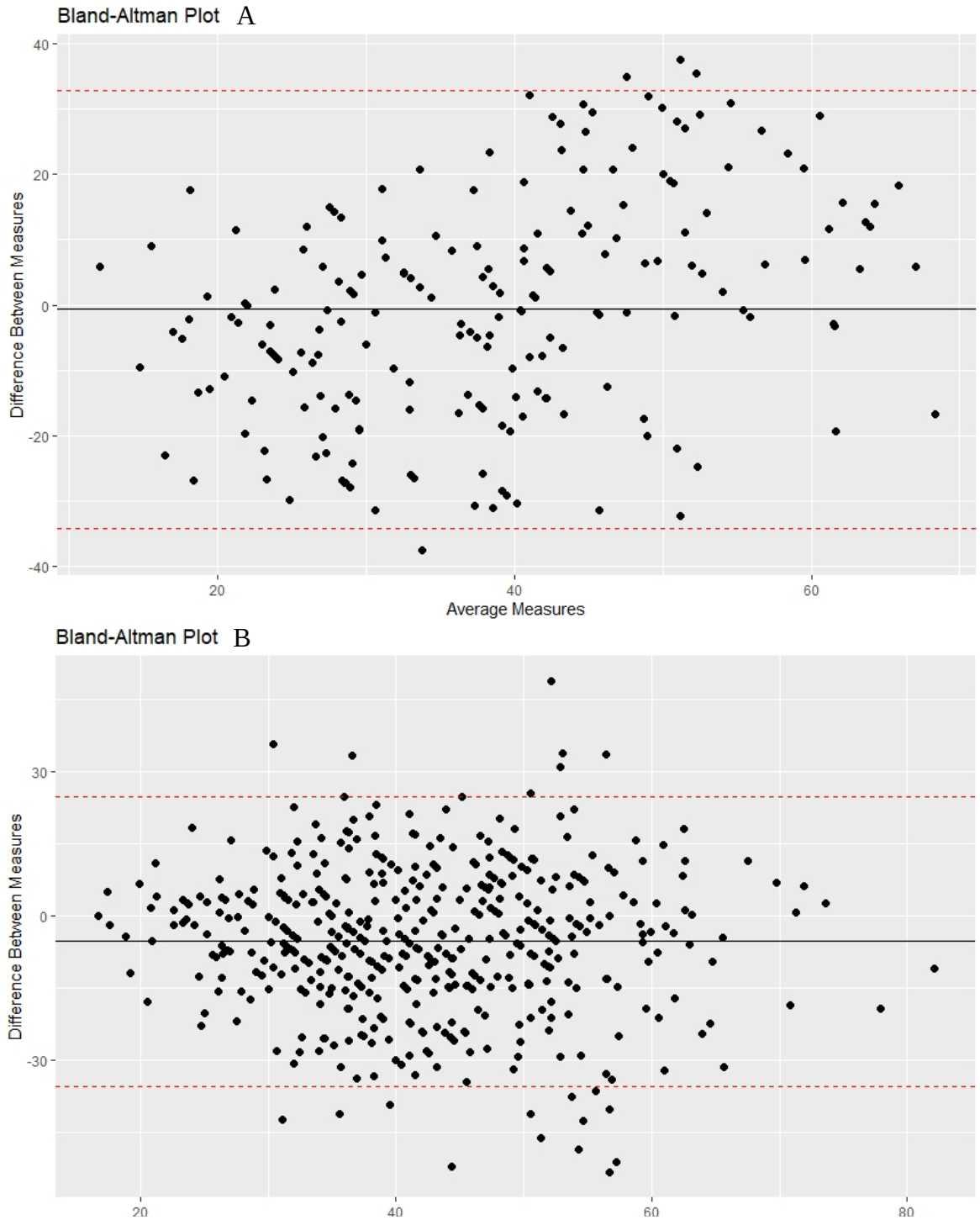


Figure 3.20 Relationship and bias between histopathological and DL determined TSP. *Bland Altman plots demonstrating the relationship and bias between histopathological TSP scores and DL algorithm generated TSP scores for the Glasgow Development Cohort (A) and Glasgow Validation Cohort (B).*

3.3.5 Validation of Deep Learning TSP Algorithm in a Clinical Trial Cohort

Given the improvement in performance of the deep learning algorithm over the machine learning algorithm in translating automated TSP scoring between independent patient cohorts, further validation of this algorithm was conducted in a substantially larger, clinical trial cohort of patients. A total of 2913 patients from the SCOT international, randomised, phase 3, non-inferiority trial (ISRCTN59757862) were selected to create a translational research cohort from the trial (TransSCOT cohort). Patients admitted to the trial had either high-risk Stage II or Stage III CRC and tissue was obtained from patients undergoing surgical resection between 2008-2013 within the UK prior to commencement of the trial. Patients were predominantly male (61%) with T – Stage III (58.2%) N – Stage 2 (57.1%) colonic disease (82.5%). No clinical exclusion criteria were applied to this cohort, however, 13 patients died within 30 days of surgery and 12 patients had missing tissue or tissue inappropriate for image analysis in this study, meaning 2887 patients were included in the survival analysis. The primary clinical outcome was Disease Free Survival measured from the date of randomization in the trial to the date of recurrence or death from any cause. Clinicopathological features associated with disease free survival were T – Stage ($P < 0.001$), N – Stage ($P < 0.001$), and tumour site ($P < 0.001$) (Table 3.5).

Clinicopathological Characteristic	N (%)	Clinical Outcome Significance
Sex		
Female	1135 (39.0)	*0.598*
Male	1778 (61.0)	
T Stage		
I	78 (2.7)	
II	250 (8.6)	*<0.001*
III	1696 (58.2)	
IV	889 (30.5)	
N Stage		
0	556 (19.1)	
1	1663 (57.1)	*<0.001*
2	694 (23.8)	
Tumour Site		
Colon	2402 (82.5)	*<0.001*
Rectum	511 (17.5)	
Chemotherapy Type		
FOLFOX	846 (29.0)	*0.322*
CAPOX	2067 (71.0)	

Table 3.5 Clinicopathological Characteristics of TransSCOT Clinical Trial Cohort. *Number (and %) of patients with clinicopathological features in TransSCOT Clinical Trial Cohort and association with Disease Free Survival*.

Following the same analysis process as in the previous two validation cohorts, the TSP scores generated by the algorithm for each patient in this cohort underwent ROC analysis to determine the predictive capability given a substantial increase in patient numbers. ROC curve analysis against the high vs low histopathological TSP patients generated an AUC = 0.828 (Figure 3.21).

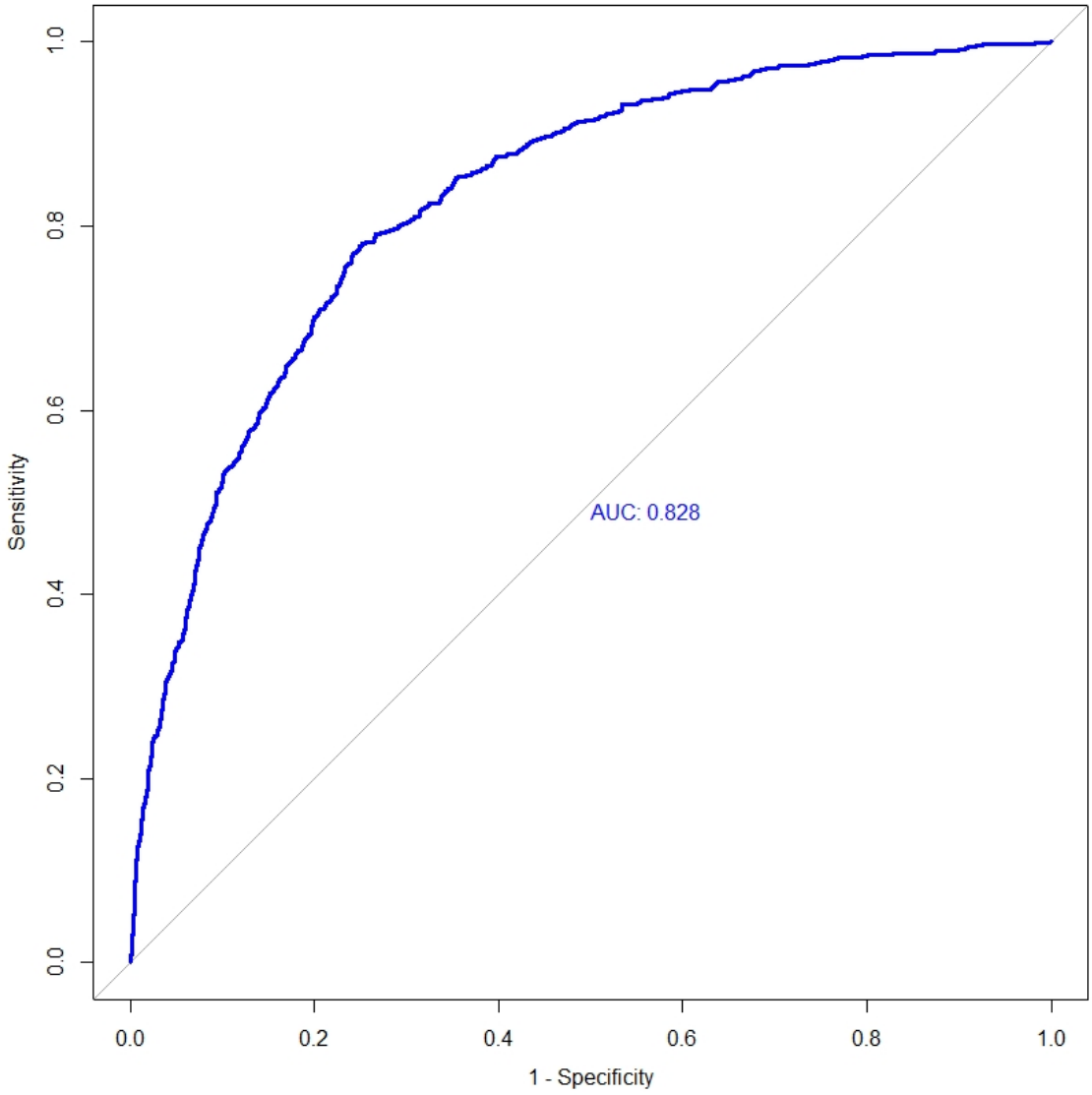


Figure 3.21 Receiver Operator Characteristic of DL algorithm generated TSP in TransSCOT Clinical Trial Cohort. *ROC curve of DL algorithm generated TSP scores against high vs low histopathological assessment in the TransSCOT cohort, AUC = 0.828.*

Survival characteristics of the DL generated TSP scores were compared to manual histopathological assessment via Kaplan – Meier analysis. The algorithm generated TSP was dichotomised into high and low stroma using the same 40% cut off generated in the Norway validation cohort using disease free survival as the primary outcome measure. Consistent with the previous validation cohorts, the DL generated high vs low stroma patient stratification showed a statistically significant association with disease free survival in the clinical trial cohort (HR 1.729 95% CI 1.474 – 2.029, P <0.0001, Figure 3.22). Cohen's Kappa metric was utilised to assess the level of statistical agreement, showing moderate concordance between the two scores, $\kappa = 0.452$ (Figure 3.22).

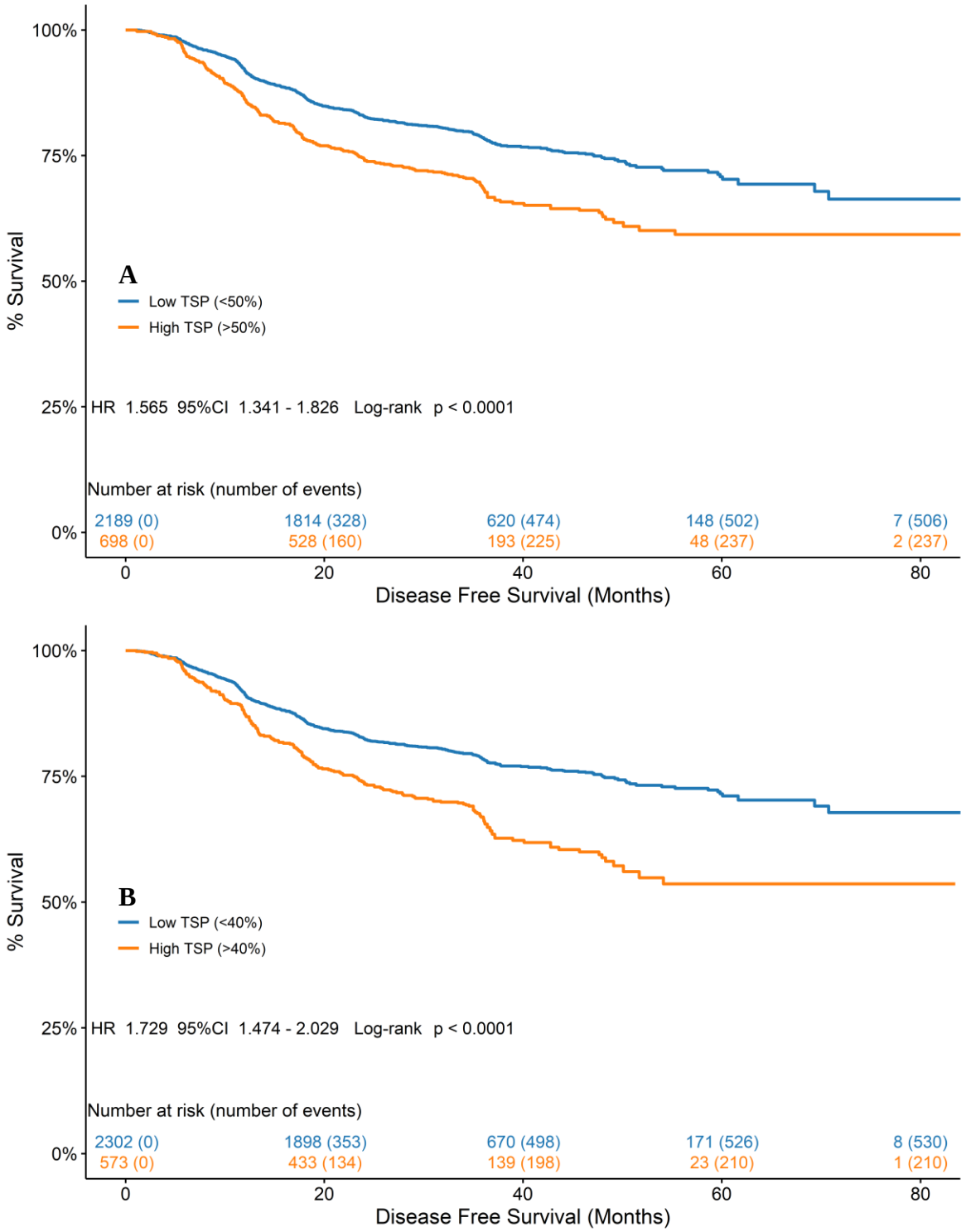


Figure 3.22 Relationship between DFS and TSP in TransSCOT Clinical Trial Cohort. *Association of DFS and TSP determined by (A) histopathological and (B) automated DL based assessment (HR 1.729 95% CI 1.474 – 2.029, P <0.0001) in TransSCOT Clinical Trial cohort. Cohen's Kappa agreement, κ = 0.452.*

Bland – Altman plots were constructed to visualise any bias and assess agreement between the manual histopathological TSP and the DL generated scores in the clinical trial cohort (Figure 3.23).

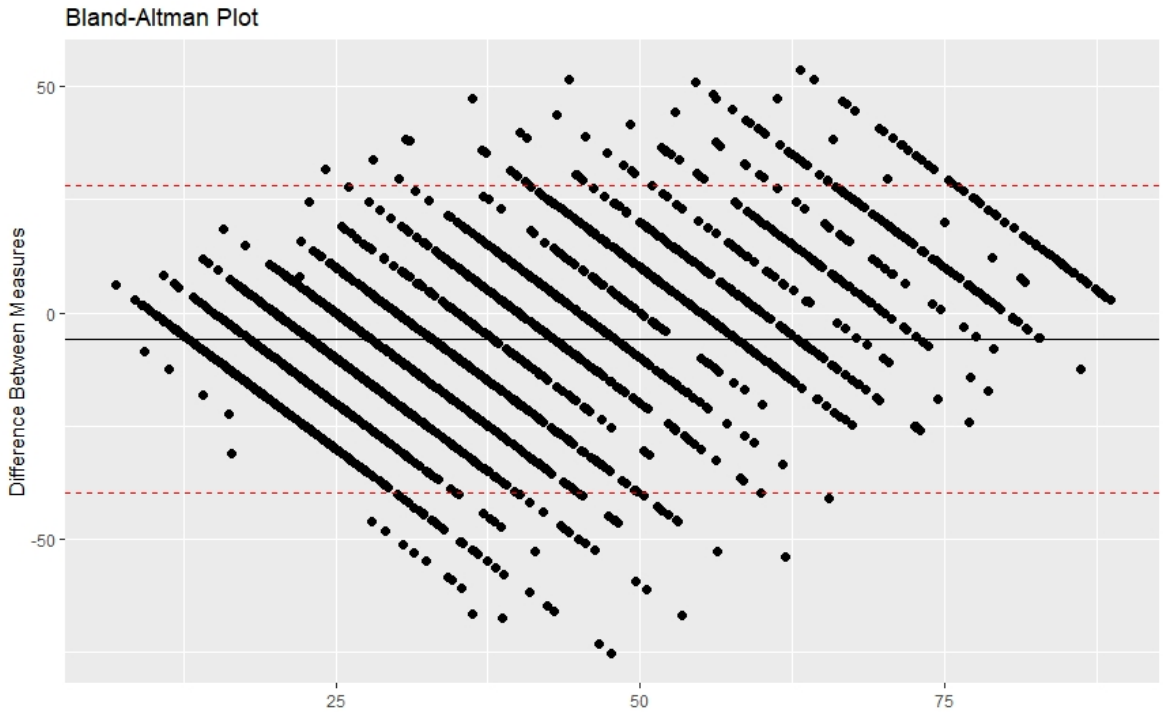


Figure 3.23 Bland – Altman Plot of Histopathological and DL Generated TSP Scores in TransSCOT Cohort. *Plots showing difference between measures and mean of measures for histopathological TSP and DL generated TSP scores for TransSCOT Clinical Trial Cohort.*

Univariate and multivariate cox regression was performed to determine the relationship between the deep learning TSP assessment, clinicopathological features and disease-free survival (Table 3.6). On multivariate analysis, the TSP stratification produced by the deep learning algorithm was significantly associated with DFS (HR 1.359 95%CI 1.155-1.599, $p < 0.001$), independent of T-Stage ($p < 0.001$), N-Stage ($p < 0.001$), and tumour site ($p = 0.031$).

Univariate				Multivariate			
	HR	95% CI	p-value		HR	95% CI	p-value
Sex							
Female	1.0			-	-	-	-
Male	1.040	0.898-1.205	0.599	-	-	-	-
T Stage				**T Stage**			
1	1.0			1	1.0		
2	1.013	0.498-2.060	0.973	2	1.116	0.548-2.271	0.762
3	1.85	0.988-3.467	0.055	3	1.9	1.012-3.564	**0.046**
4	3.217	1.715-6.037	**<0.001**	4	3.275	1.737-6.178	**<0.001**
N Stage				**N Stage**			
0	1.0			0	1.0		
1	1.436	1.155-1.787	**0.001**	1	1.795	1.436-2.244	**<0.001**
2	2.821	2.249-3.538	**<0.001**	2	3.032	2.411-3.814	**<0.001**
Site				**Site**			
Left	1.0			Left	1.0		
Right	0.687	0.557-0.848	**<0.001**	Right	0.788	0.634-0.979	**0.031**
Treatment							
FOLFOX	1.0			-	-	-	-
CAPOX	1.084	0.924-1.272	0.319	-	-	-	-
TSP				**TSP**			
Low	1.0			Low	1.0		
High	1.646	1.405-1.929	**<0.001**	High	1.359	1.155-1.599	**<0.001**

Table 3.6 Relationship between TSP, Clinicopathological Features and Disease-Free Survival in the TransSCOT Cohort. Hazard ratios and 95% confidence intervals calculated via Cox proportional hazards regression and multivariate analysis conducted using backwards conditional method.

3.4 Discussion

The TNM staging criteria of colorectal cancer remains the gold standard for clinical prognosis and patient stratification across the world. However, it has been shown that there is notable variation in survival outcomes for patients of a seemingly similar disease state, particularly in stage II disease where 5 – year survival outcomes for stage IIa vs stage IIb are 84.7% and 72.2%, respectively (O'Connell et al., 2004). This is likely due to the heterogeneous nature of the CRC microenvironment and the complex mechanisms underpinning its progression, indicators of which are conspicuously lacking in the TNM staging criteria. This has directed research to investigate readily assessable components of the TME that hold prognostic significance and could be translated to routine diagnostic pathology. One of these metrics is the Tumour Stroma Percentage (TSP) or Tumour Stroma Ratio (TSR), a semi – quantitative assessment of the relative stromal component of the tumour area that is assessable on a single H&E section generated during routine diagnosis. The TSP has demonstrated prognostic significance in a large number of studies across multiple stages of disease (Huijbers, 2013; Huijbers et al., 2013; Park et al., 2014b; Park, McMillan, et al., 2016; van Pelt et al., 2018b), however, it has seen little clinical implementation due to the inconsistency of measures of intra and interobserver agreement. Image analysis of H&E WSI presents possible solution to this issue by having a single, consistent observer conduct the assessment. In the present study, image analysis – based approaches to conduct TSP assessment on H&E WSI were evaluated in multiple, distinct cohorts of CRC patients to formulate a method that could reliably stratify patients for survival.

Given the intrinsic ability of deep learning algorithms to adapt to novel data, these approaches are an obvious first choice for conducting high level tissue classification on heterogeneous data populations such as clinical H&Es. However, there is a significant cost associated with performing DL – based studies, notably high – performance graphics processing units that are required to handle convolutional neural network training and application, that presents a roadblock to use in developing nations and rural areas without substantial IT infrastructure. To address this, an initial approach was taken to develop an image analysis method for TSP assessment using traditional machine learning methods. The aim was to use a task specific approach whereby the decision forest algorithm was trained on manually crafted features (Figure 3.1) that provided high – level information on the relationship between the tumour and stroma that RGB colour space does not. When this algorithm was applied to the Norway cohort of patients, from which the training data set was constructed, it was able to significantly stratify patients for CSS based on a cut off value of

40% stromal component, performing comparably to manual histopathological assessment on the same sections (Figure 3.7). However, as is generally the case when applying machine learning – based image analysis algorithms to novel image data, the algorithm stratification did not show a significant association with CSS when translated to the Glasgow Development and Validation cohorts (Figure 3.10 & Figure 3.11). There are numerous possible causes for this inability to translate ML algorithms to new image data, within this study specifically it likely due to the use of non – standardized training data but could arise from lack of section QC, scanner age and type, or histological processing. The use of Training Data Set 2 (2.4.2.2) proved too computationally burdensome for the CPU – based ML algorithm, and while attempts were made to ameliorate the effect that using WSI from the same cohort through targeted section selection accounting for histological variance, this was evidently insufficient, and a larger training data set along with both the training and test data to undergo stain normalization is likely required (Bejnordi et al., 2016).

Therefore, an alternative approach was adopted. Convolutional neural network – based approaches are a logical solution to the issues in translatability discussed above. To apply deep learning to this task, a UNET CNN trained on the colour – normalized Training Data Set 2 (2.4.2.2) for 9 tissue classes was used to segment the WSI and quantify the TSP for each patient. While studies utilising deep learning to classify the type of desmoplastic stroma and assess stromal cellular content have shown promise in regards to prognosis, the TSP is the most readily translatable stromal assessment to clinical practice and has been extensively validated in multiple cohorts, therefore, this was chosen as the most appropriate application of deep learning for this study (I. P. Nearchou et al., 2021). Initial validation in the Norway cohort demonstrated an improved prognostic stratification over the ML algorithm and outperformed manual histopathological assessment for patient survival (Figure 3.16). When applied to the Glasgow Development and Validation cohorts, the DL algorithm demonstrated the translational capacity lacking in the ML algorithms performance, significantly stratifying patients for cancer specific survival in both cohorts (Figure 3.18 & Figure 3.19), and outperforming manual histopathological assessment in the Glasgow Development cohort (Figure 3.18). This translational ability was further demonstrated when the algorithm was applied to the much larger TransSCOT Clinical Trial cohort; in 2875 stage II – III patients, the DL algorithm significantly stratified patients for disease free survival as well as again outperforming manual histopathological assessment (Figure 3.22).

A notable limitation of the present study is the inconsistency in statistical measure of agreement through Cohen's Kappa analysis. The ML algorithm demonstrated moderate

correlation with manual histopathological assessment ($\kappa = 0.443$) in the Norway cohort, from which the training data was derived, but this expectantly dropped when translated to novel cohorts ($\kappa = 0.151$ & 0.11). The DL algorithm showed moderate agreement with manual histopathological assessment in the Norway and TransSCOT cohorts ($\kappa = 0.471$ & 0.452, respectively) but the ML algorithm did not perform to the same level in the Glasgow Development and Validation cohorts showing fair and slight concordance ($\kappa = 0.21$ & 0.18, respectively) in those cohorts. It is noteworthy that while subjective interpretation of the Cohen's Kappa thresholds would deem the algorithm performance in the Norway and TransSCOT cohorts as somewhat poor for a clinical assessment, there was still >80% concordance between the algorithm and manual histopathological assessment in both cohorts and all correlative statistical measures are substantially significant.

The lower Kappa values in the two Glasgow cohorts is likely explainable through two hypotheses. Efficacious image processing of clinical WSI is contingent on slides being of good enough scanning and histological quality for image analysis to be appropriate, which, given the age of the sections used in these cohorts, the date they were scanned and the type / age of the scanner used, and the fact that physical annotations made by a pathologist on the Glasgow Validation cohort section coverslips caused focussing issues during scanning, likely contributed to sub – optimal tissue classification on these cohorts' WSI. The Norway and TransSCOT sections were stained and scanned much more recently lending credence to the hypothesis that adequate section QC prior to image analysis is a necessity for ensuring validity of the clinical data extracted from the WSI. Secondarily, there were multiple sections available for each patient in the Glasgow Validation cohort ($N = 1 – 6$) but it was unknown which section was used for the original manual histopathological assessment, meaning that taking the average of N sections per patient is likely the reason for the statistical agreement being lowest in this cohort.

In summary, this chapter clearly demonstrates the efficacy of deep learning – based image analysis algorithms in conducting automated TSP assessment across multiple, geographical and process independent clinical cohorts, often outperforming manual histopathological assessment. Given the prognostic significance of assessing the tumour associated stroma, algorithms such as this could provide additional key information to pathologists for directing clinical decisions, and through this workflow, be able to assess algorithm performance on a case – by – case basis via visualization of the algorithm overlay for the specific WSI. Future work should be directed towards optimizing algorithms such as this for clinical application and determining the extent to which statistical measures of

agreement with manual histopathological assessment are used as the validatory endpoint for these studies when the aim is to address interobserver variability in the assessment.

4. Quantitative Assessment of the Local Immune Response to Colorectal Cancer by Image Analysis

4.1 Introduction

One of the most well characterised, prognostic features of the CRC tumour microenvironment is the local host inflammatory response, with the survival benefit of a prominent lymphocytic infiltrate in gastric tumours having been first described in 1931 (Maccarty, 1931b). Following this, a series of studies beginning in 1967 further demonstrated the significant improvement in prognosis conferred by a prominent lymphocyte response, both peritumourally and intratumourally, with Jass and colleagues identifying the response at the invasive margin as a stage independent factor in 1987 (Jass et al., 1987; C. S. D. Roxburgh & McMillan, 2012).

These early studies utilised diagnostic H&E sections taken during routine pathology and characterised the generalised inflammatory response with a few studies investigating lymphocyte and plasma cell populations specifically (de Mascarel et al., 1981; Zamcheck et al., 1975). The advent of immunohistochemistry allowed more targeted investigation into the prognostic role of specific immune cell subtypes through molecular characterization, with cancer cell nest $CD8^+$ cytotoxic T – lymphocytes soon being identified as another stage independent prognostic factor of the localised immune response (Naito et al., 1998). Subsequent studies utilizing IHC, confirmed initial interest in lymphocytes as the key immune cell type modulating survival through investigation into $CD3^+$ cells (Nagtegaal et al., 2001), and further elucidated the prognostic role of specific lymphocyte sub populations, such as $CD45RO^+$ memory T – cells (Öberg et al., 2001). Galon and colleagues' investigation into the role of both cellular subtype and localisation within the tumour identified the presence of $CD3^+$ $CD8^+$ $CD45RO^+$ Granzyme B^+ lymphocytes associated with a T_H1 adaptive immune response exerted a significantly beneficial effect on survival. This data ultimately culminated in the introduction of the Immunoscore (Galon et al., 2006), a combined assessment of $CD8^+$ and $CD45RO^+$ (later replaced with CD3 due to background and loss of antigenicity in stored sections) at the invasive margin and within the tumour core, quantified through image analysis that is able to outperform TNM staging for prognostic utility (Galon et al., 2014).

Though molecular investigation into the prognostic role of lymphocytes has yielded significant data, there has been little clinical adoption of systems like the Immunoscore, due to the cost of routinely performing IHC as part of the diagnostic process. This has turned attention back towards earlier methods of prognostic stratification through semi – quantitative lymphocyte assessment, namely those performed on diagnostic H&E sections, such as that of Jass and colleagues, and Klintrup and Mäkinen (Jass et al., 1987; Klintrup et

al., 2005). The Jass system, originally established in rectal cancers, binarily assigns patients as possessing a high immune infiltrate (denoted conspicuous in the original study) when there is a scattering of lymphocytes within a distinctive connective tissue mantle at the invasive margin and low otherwise. The Klintrup – Mäkinen system expands upon Jass' work by stratifying patients into four distinct groups based on the absence (0), patchy distribution (1), band – like reaction (2), and florid cup – like reaction (3) of lymphocytes at the invasive margin but is dichotomized into high (3/2) and low (1/0) by convention. While both systems hold significant prognostic value and have been validated in a number of studies, neither have been adopted into clinical practice due to notable issues with inter – and intra – observer agreement and the lack of a standardized assessment criteria (Hynes et al., 2017b).

Image analysis, particularly with the growing application of deep learning, has begun to yield promising data that could bridge the gap between quantitative IHC and semi – quantitative H&E assessment of the inflammatory response. Prediction of survival by neural networks through unsupervised and semi – supervised training is a well – established method for deriving clinical information from routinely produced H&Es (Bychkov et al., 2018b; Campanella et al., 2019b). However, recent studies have utilized neural networks and traditional image processing to look to quantify both histological features and cell types in order to reproducibly obtain more pertinent data in the setting of CRC. Pai and colleagues demonstrated quantitative differences in the amount of inflammatory stroma and TILs between pMMR and dMMR tumours following global tissue segmentation on WSI using commercially available, neural network coupled digital pathology software (Pai et al., 2021). Using non – deep learning image analysis in the open source QuPath software (Bankhead et al., 2017), Väyrynen and colleagues quantitatively assessed the densities of stromal lymphocytes, plasma cells, neutrophils, and eosinophils in H&E TMAs showing all four cell types to be significantly associated with CSS (Vayrynen et al., 2020).

The aim of the present study, therefore, is to assess combination image analysis techniques to translate the Klintrup – Mäkinen grading system to a quantitative, automated approach. Additionally, to investigate different image analysis algorithms to reliably and reproducibly automate the quantitative scoring of lymphocytes on routine clinical H&E WSI and compare these methods to histopathological grading in terms of patient stratification for disease specific survival.

4.2 Materials and Methods

4.2.1 Histopathological Assessment

Manual histopathological assessment of the Klintrup – Makinen grading criteria was carried out on H&E slides / WSI as described in 2.3.1 for each patient cohort using the criteria described in the original paper (Klintrup et al., 2005). Briefly, the lymphocyte density was assessed at the invasive margin of the tumour on a section taken from the deepest point of invasion using a four – point scale: 0 – No evidence of inflammatory response at the invasive margin, 1 – Patchy distribution of lymphocytes along the invasive margin, 2 – inflammatory response forming a band – like structure across a significant portion of the invasive margin, 3 – Prominent inflammatory response forming a florid cup – like structure across the invasive margin with evidence of tumour island destruction. Subsequently, as is done by convention, the four – point score was dichotomized with KM grades 0/1 being classed as "Weak" inflammatory response and KM grades 2/3 being classed as a "Strong" inflammatory response. All Klintrup – Mäkinen grading in this study was conducted by an independent observer not involved with image analysis, names of observers stated where relevant, and grades were obtained from existing databases.

4.2.2 Statistical Analysis

Patient cohort clinicopathological characteristics and test of association with CSS / DFS were determined in IBM SPSS version 25 (IBM, New York, USA). P values of <0.05 were considered to demonstrate a statistically significant association between clinicopathological features and clinical outcomes.

All other statistical analyses for this study were conducted in RStudio (RStudio, Boston, MA). Optimal cut off values for immune infiltrate scores were determined by using CSS / DFS as the endpoint with the maxstat and survminer packages. Receiver operator characteristic curves and associated AUC were plotted and calculated using the pROC package. Survival analysis was conducted and associated Kaplan – Meier survival curves produced using the survival, survminer, and survMisc packages, with the log rank statistic used to assess association between immune grade stratification and CSS / DFS. Hazard ratios and associated 95% confidence intervals for survival analysis were calculated using the Cox proportional hazard model.

4.3 Results

4.3.1 Development of Image Processing Algorithms to Automatically Detect Lymphocytes on Clinical H&E Sections

Automated detection of lymphocytes was conducted on H&E sections utilizing a threshold – based image processing algorithm, initially designed to be a pan – nucleus detector with additional processing steps to exclude non – lymphocyte cell types. To segment nuclei from other tissue morphologies, images in RGB colour space were colour deconvolved into the component stain layers of Haematoxylin and Eosin. From which an 11 x 11 kernel, second order / 5 x 5 kernel second order Laplacian of Gaussian (LoG) filter was applied to the Haematoxylin colour deconvolved layer with a threshold value of >20 applied to the resultant greyscale feature vector to isolate Haematoxylin dense areas representing nuclei. The areas detected within this greyscale feature were used to generate centroids for each nucleus via a local intensity maxima marking function. Nuclear boundaries were generated from the nuclei centroids with an inferred maximum membrane distance of 4µm using the background from the initial LoG classification as a delimiter. Finally, a hole filling function was applied within the outlined nuclei to create a mask encompassing the nucleus of all cell types on the section.

To isolate lymphocytes from other cell type nuclei, a series of size, shape, and stain intensity thresholds were applied to the nuclear mask. Large nuclei and small pixel clusters not belonging to nuclei were removed using a size exclusion threshold of <10 / >50µm^2. Elongated nuclei were excluded by applying an isoperimetric quotient, given by $\frac{4 \cdot \pi \cdot Area}{Perimeter^2}$, threshold of <0.7 to the nuclear mask, given the isoperimetric quotient of a perfect circle is 1. The resultant classification consisted of lymphocytes and eosinophilic immune cells with haematoxylin dense nuclei; therefore, to isolate lymphocytes from other immune cell types, cells were excluded based the object eosin content by applying an exclusion intensity threshold of <150 to a 5 x 5 kernel, maximum filtered eosin colour deconvolved layer encompassing >50% of the nuclear object. Representative images are shown in Figure 4.1.

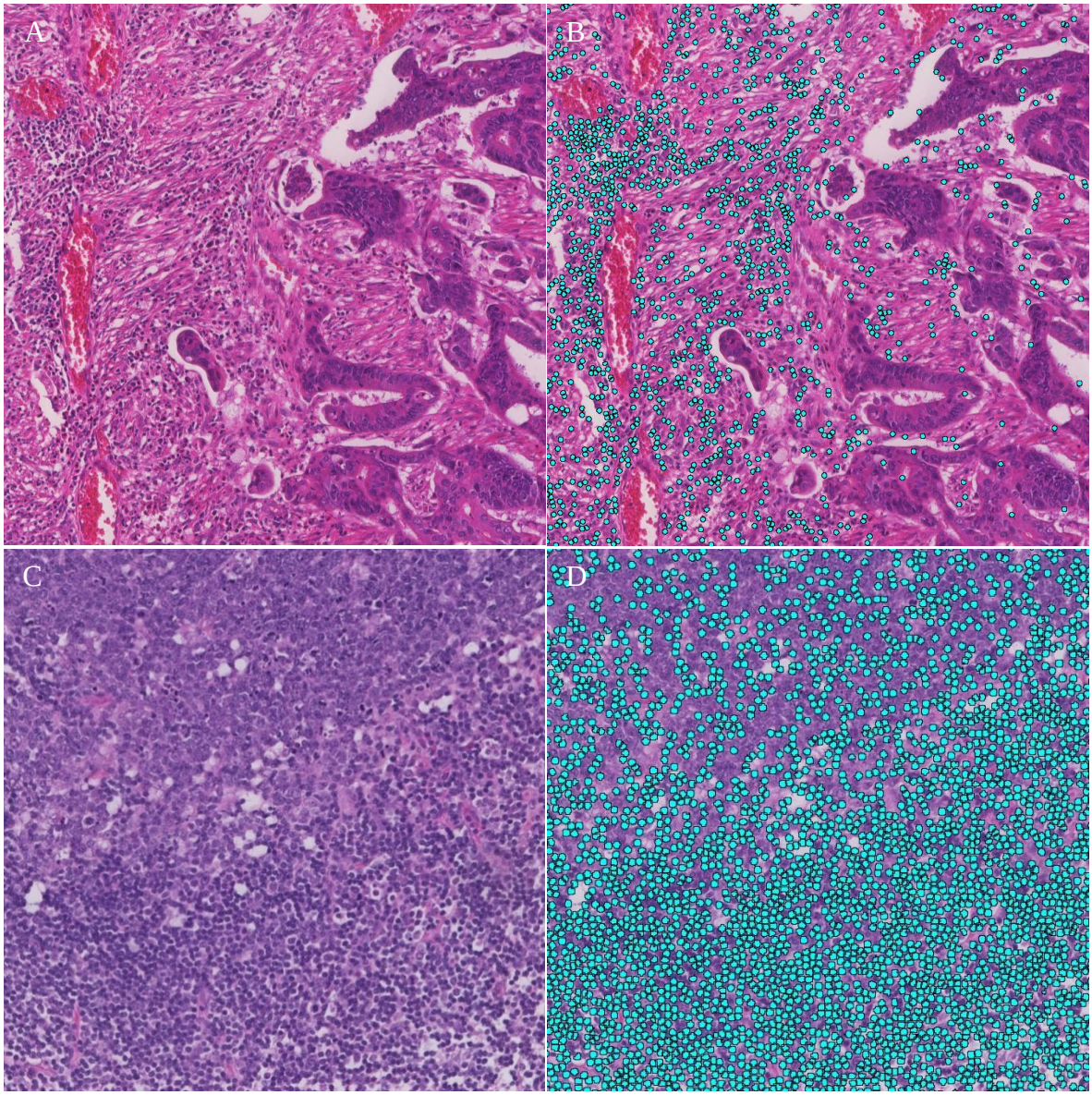


Figure 4.1 Lymphocyte Detection Algorithm Peformance. *Images demonstrating lymphocyte detection algorithm classification at invasive margin (AB) and edge of germinal center of lymph node (CD) taken from a CRC H&E section. Detected lymphocytes are shown in blue. 10x magnification.*

The lack of a standardised training data set of annotated H&E sections for lymphocytes necessitated the lymphocyte detection algorithm be developed through traditional image processing methodologies. Therefore, all threshold values described above were determined via visual assessment and validation across multiple sections throughout the development process to ensure accuracy and utility across histologically variant sections. To validate the accuracy of the threshold values chosen, two further algorithms were adapted from this initial lymphocyte detection algorithm with changes made to specific haematoxylin and eosin – based threshold values to alter the number and potential type of cells detected. The first algorithm relaxed the threshold value of the LoG filtered haematoxylin colour deconvolved layer to 15, allowing less haematoxylin dense nuclei to be detected, and the threshold value of the maximum filtered eosin colour deconvolved layer to 125 encompassing >25% of the cell object. The second algorithm increased the threshold value of the LoG filtered haematoxylin colour deconvolved layer to 30, meaning only very haematoxylin dense nuclei were detected, and threshold value of the maximum filtered eosin colour deconvolved layer to 200 encompassing >50% of the cell object, aiming to ensure exclusion of all eosinophilic immune cells but potentially at the cost of detecting all lymphocytes.

All three algorithms underwent a two – step validation process, first to quantitatively assess the accuracy of the lymphocyte detection against sections stained via IHC for the pan – T lymphocyte marker CD3 via Dice Similarity Coefficient (DSC) analysis (Dice, 1945), and to compare the prognostic performance of each algorithm against the patient stratification produced by Klintrup – Mäkinen grading via Kaplan – Meier survival analysis. This was done based on the hypothesis that, given the conflicting prognostic effect of different infiltrating immune cell types, increasing the number and type of cells detected by the algorithm would be accompanied by a drop in prognostic significance whereas decreasing the number but potentially increasing the accuracy of lymphocyte detection would produce a similar or increased significance in prognosis compared to the initial algorithm. Additionally, DSC values could theoretically decrease with stricter cell segmentation thresholds or increase with more relaxed cell segmentation thresholds, therefore, the aim of the two – step validation process was to find the optimum balance between prognostic significance and cell detection accuracy, from which the most appropriate algorithm would be determined and applied to further patient cohorts.

To assess the accuracy of the lymphocyte detection algorithms via DSC analysis, 20 FFPE blocks were taken at random from either the Glasgow Development or Glasgow Validation cohorts by an independent researcher (Aula Ammar), and two sequentially cut sections were stained for H&E and CD3. Both sections were scanned at 20x objective magnification and the WSI aligned using the "Tissuealign" module in the Visiopharm software with the automatic alignment function, following which the tissue detection algorithm (detailed in 2.4.3.1) was applied to the H&E section, to remove slide glass background, and overlayed on to the aligned CD3 section. Cells were segmented and classified as positive or negative on the CD3 IHC sections via the same algorithm used to perform Ki67 %PI analysis described in 2.4.6.2 and later in 5.3.1, and DSC scores were generated by applying each lymphocyte detection algorithm to the respective overlayed H&E section. The high threshold algorithm showed the greatest accuracy of lymphocyte detection with a DSC of 0.67, with the low threshold algorithm showing the poorest cell detection accuracy (DSC = 0.47) and the initial algorithm showing intermediate accuracy (DSC = 0.58) (Figure 4.2).

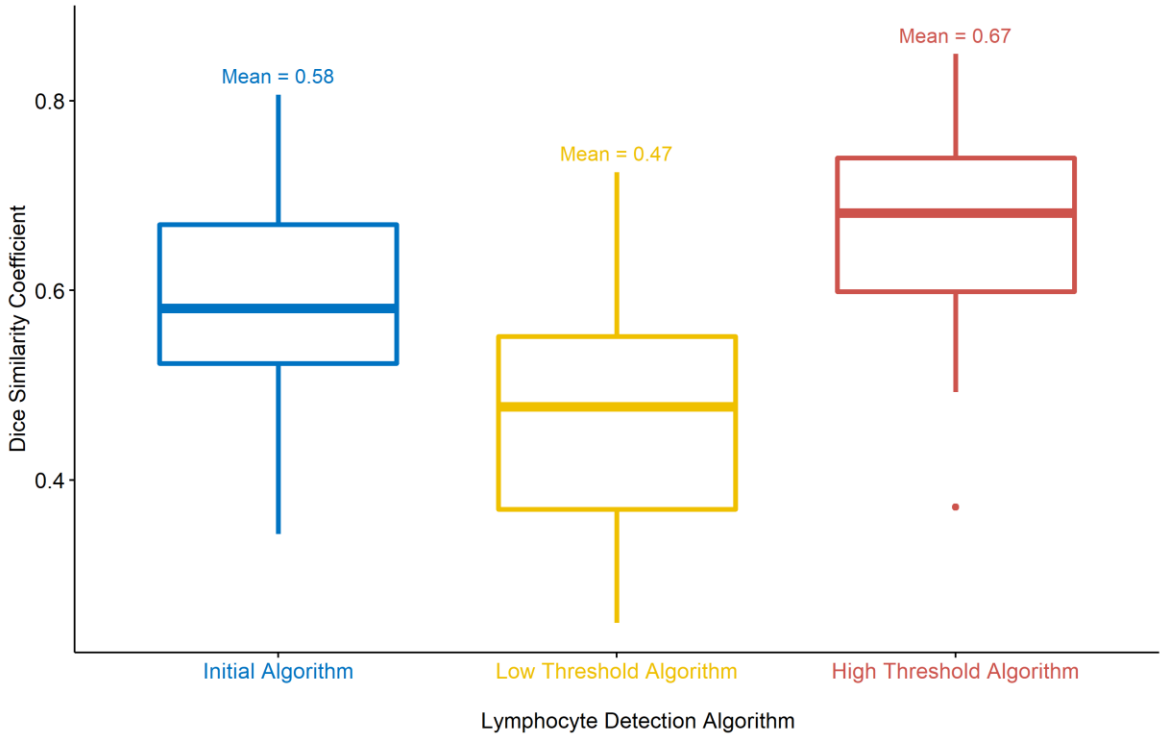


Figure 4.2 Distribution of Dice Similarity Coefficients of Lymphocyte Detection Algorithms. *Distribution and means of DSC scores generated for each H&E lymphocyte detection algorithm against sequentially cut CD3 IHC sections.*

In the second validation step, Kaplan – Meier survival analysis was used to compare the prognostic significance of each lymphocyte detection algorithm to the patient stratification produced by KM grading. Initially, each H&E section from the Glasgow Development Cohort (N = 226) was roughly annotated for the invasive margin by the researcher conducting the image analysis (CB) to produce an ROI for analysis, from which the lymphocyte densities (N cells / mm^2) outputted by each detection algoirthm were determined. To assess the assocation of the image analysis algorithms with manual histopathological assessment, ROC analysis was conducted using the lymphocyte densities from each algorithm plotted against the patient stratification produced by KM grading. Following the same trend as the detection accuracy analysis, the high threshold algorithm showed the best association with KM grading (AUC = 0.771), the low threshold algorithm the poorest association (AUC = 0.762), and the initial algorithm intermediate association (0.765) (Figure 4.3).

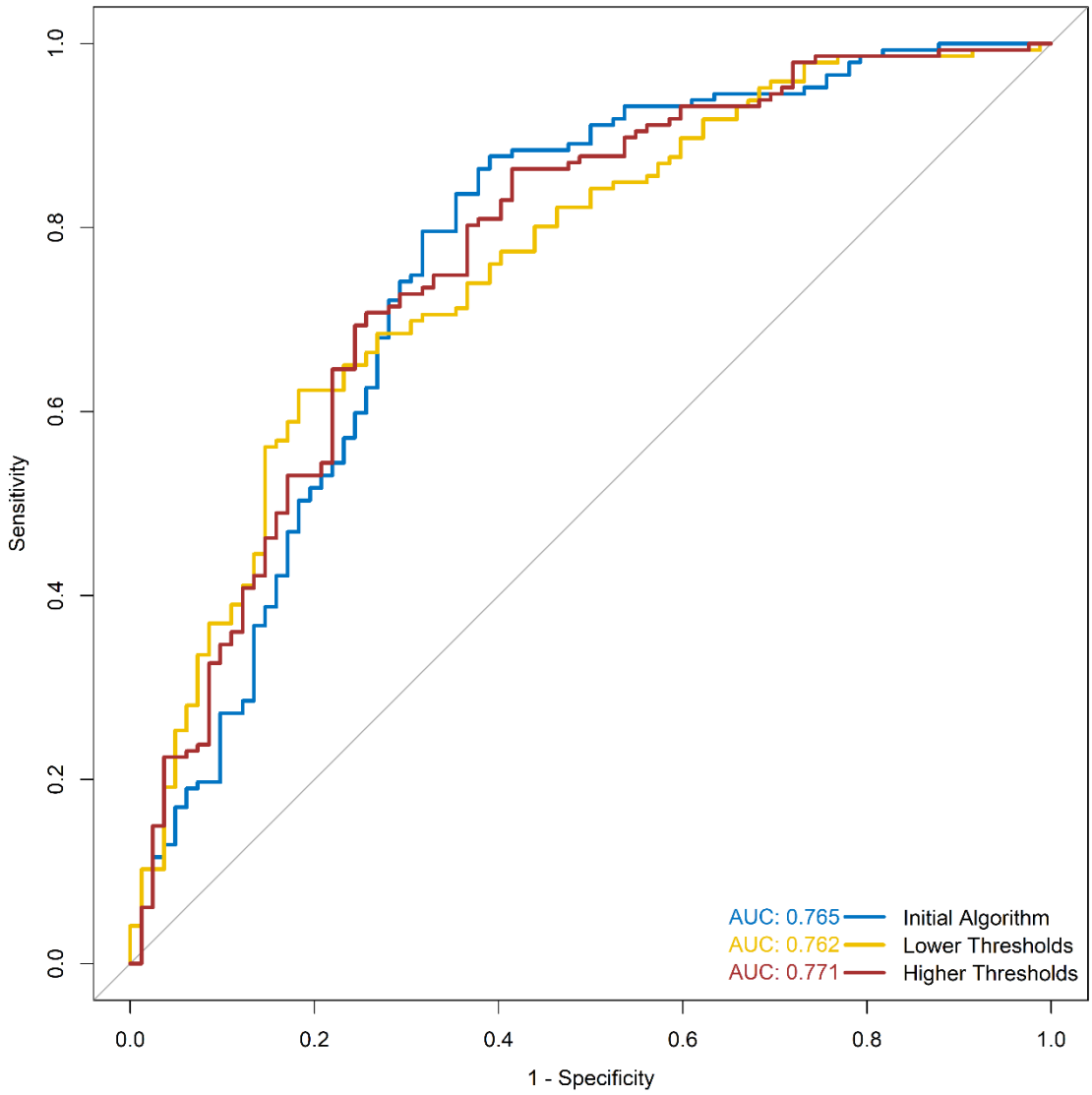


Figure 4.3 Receiver Operator Characteristic of Image Analysis Determined Lymphocyte Densities. *Lymphocyte densities (cells / mm^2) determined by three image analysis lymphocyte detection algorithms compared to KM grading stratification in the Glasgow Development Cohort.*

To compare prognostic stratification by the algorithm determined cell densities to that of manual histopathological KM grading, optimal cut off values based on CSS were determined for each algorithm using the maxstat and survminer packages in R, which were used to stratify patients into high vs low inflammatory infiltrate for each algorithm. The cut off value determined for the initial, low threshold, and high threshold algorithm was 2.25, 7.30, and 0.88 cells / mm^2, respectively (Figure 4.4).

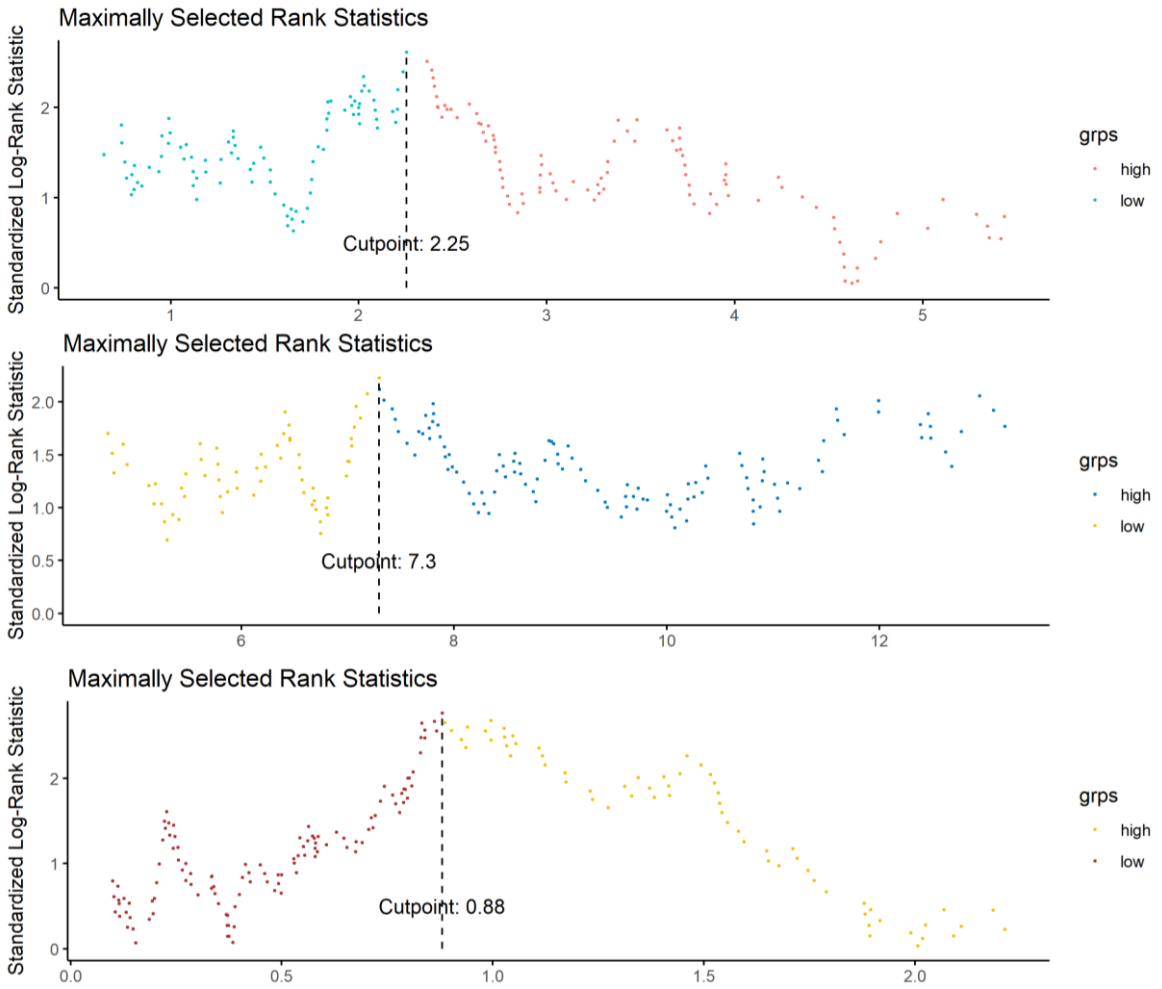
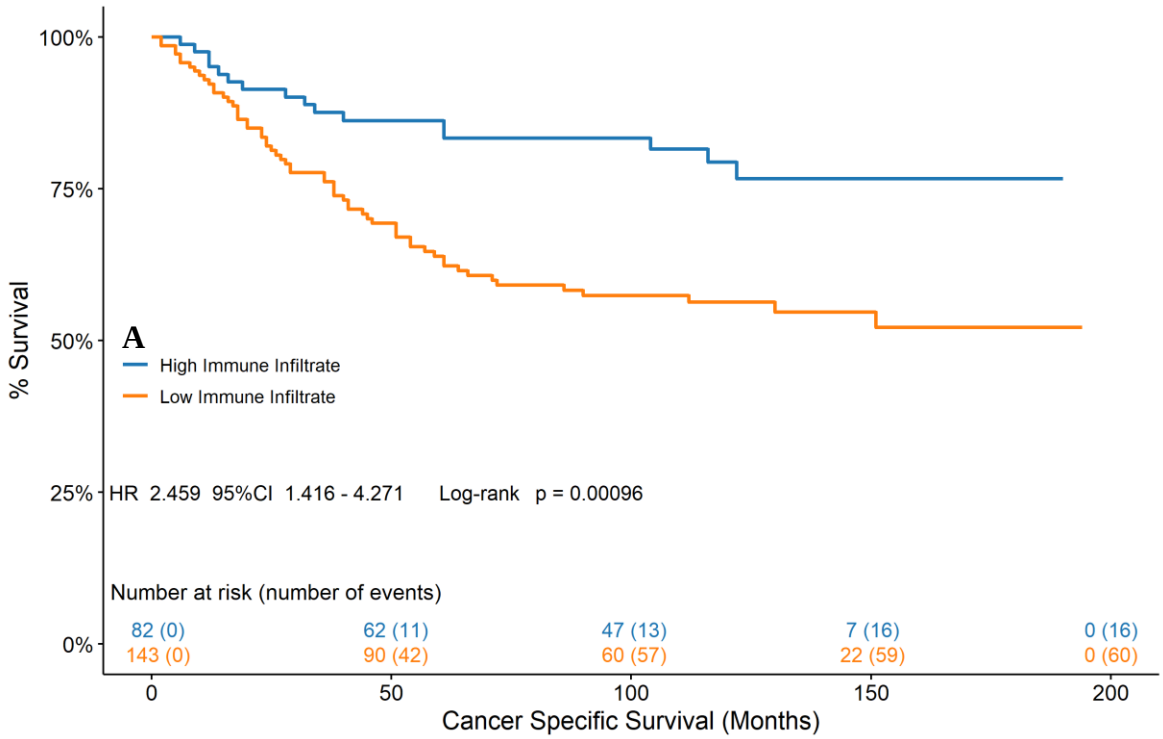
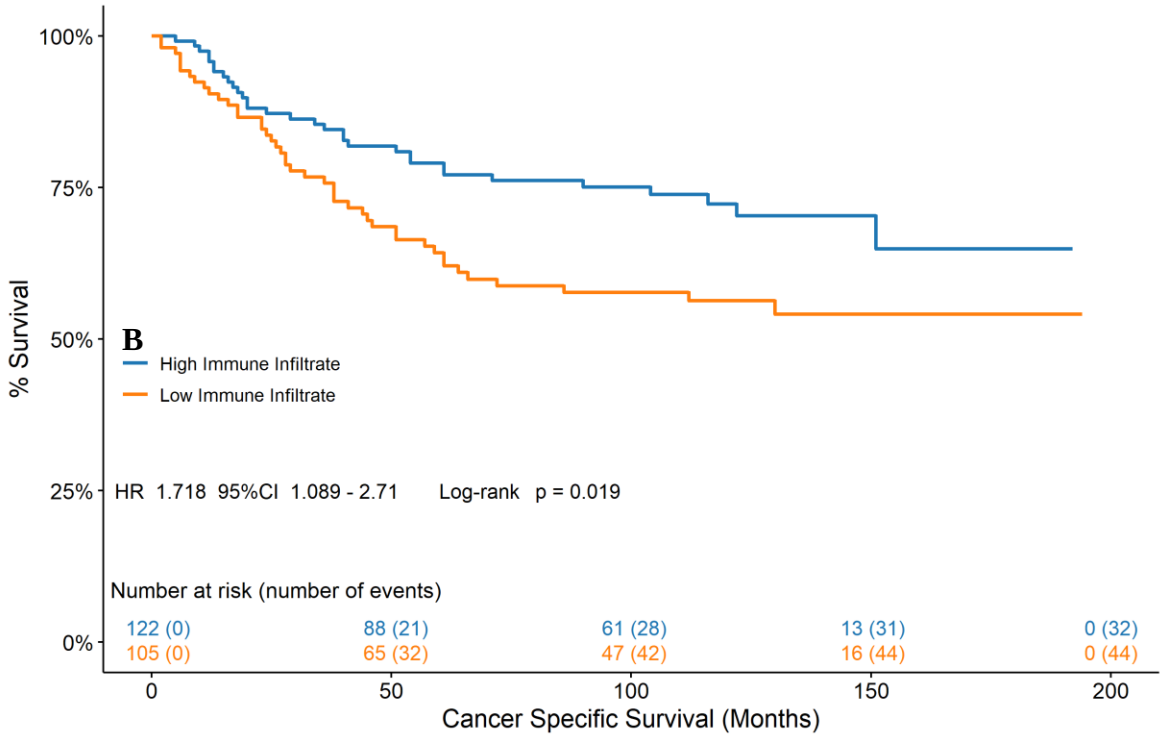


Figure 4.4 Determining Optimal Cut Off Values for Lymphocyte Detection Algorithms in Glasgow Development Cohort. *Distribution of events and optimal cut off values for lymphocyte density determined by initial (top), low threshold (middle), and high threshold (bottom) lymphocyte detection algorithms based on cancer specific survival in Glasgow Development Cohort.*

The cut off values determined via maximally ranked statistics analysis were used to stratify patients into high vs low immune infiltrate and assessed for CSS prognostic significance against manual KM grading by Kaplan – Meier survival analysis. Stratification by all three algorithms demonstrated statistically significant associations with CSS, with the high threshold algorithm producing the greatest prognostic significance and the most comparable to manual histopathological assessment, HR 2.1 95%CI 1.249 – 3.53 *P = 0.0042* and HR 2.459 95%CI 1.416 – 4.271 *P = 0.00096*, respectively. The initial lymphocyte detection algorithm demonstrated the poorest association with CSS (HR 1.718 95%CI 1.089 – 2.71 *P = 0.019*) and the low threshold algorithm showed intermediate association with CSS (HR 1.73 95%CI 1.103 – 2.714 *P = 0.016*) (Figure 4.5).


100%
75%
50%
25%
0%
% Survival
A
High Immune Infiltrate
Low Immune Infiltrate
HR 2.459 95%CI 1.416 - 4.271
Log-rank p = 0.00096
Number at risk (number of events)
82 (0)
62 (11)
47 (13)
7 (16)
0 (16)
143 (0)
90 (42)
60 (57)
22 (59)
0 (60)
0
50
100
150
200
Cancer Specific Survival (Months)


100%
75%
50%
25%
0%
% Survival
B
High Immune Infiltrate
Low Immune Infiltrate
HR 1.718 95%CI 1.089 - 2.71
Log-rank p = 0.019
Number at risk (number of events)
122 (0)
88 (21)
61 (28)
13 (31)
0 (32)
105 (0)
65 (32)
47 (42)
16 (44)
0 (44)
0
50
100
150
200
Cancer Specific Survival (Months)

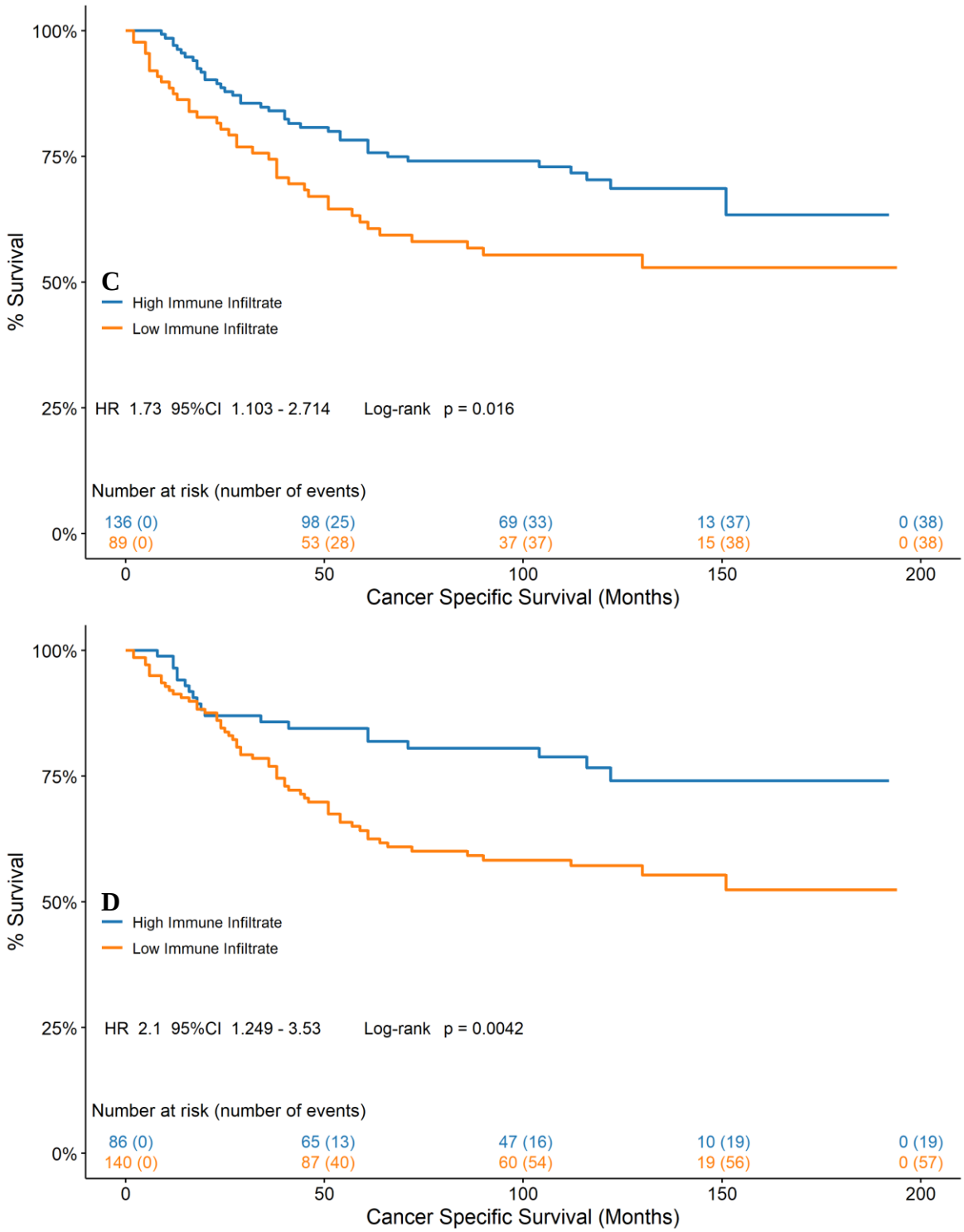


Figure 4.5 Relationship Between Lymphocyte Infiltrate and CSS in Glasgow Development Cohort. *Association of CSS and lymphocyte infiltrate assessed by histopathological Klintrup – Mäkinen grading (A), initial lymphocyte detection algorithm (B), low threshold lymphocyte detection algorithm (C), and high threshold lymphocyte detection algorithm (D) in the Glasgow Development Cohort.*

To further assess the association of the image analysis algorithms with histopathological assessment through comparative statistical agreement, Cohen's Kappa scores were generated between the stratified image analysis cell densities and the stratified high vs low KM grades for each patient. The high threshold algorithm showed the greatest statistical agreement with the histopathological assessment ($\kappa = 0.48$), the low threshold algorithm showed the poorest association ($\kappa = 0.1$), and the initial algorithm showed intermediate association ($\kappa = 0.35$) (Table 4.1).

		High KM Grade (2/3)	*Low KM Grade (0/1)*	***Cohen's Kappa***
Initial Algorithm	*High Immune Infiltrate*	64	18	**0.35**
	Low Immune Infiltrate	57	87	
Low Threshold Algorithm	*High Immune Infiltrate*	79	3	**0.1**
	Low Immune Infiltrate	122	20	
High Threshold Algorithm	*High Immune Infiltrate*	56	26	**0.48**
	Low Immune Infiltrate	29	114	

Table 4.1 Statistical Agreement of Lymphocyte Detection Algorithms and Klintrup – Mäkinen Grading. *Confusion matrices and Cohen's Kappa scores of statistical agreements between stratified image analysis lymphocyte densities and KM grade in the Glasgow Development cohort.*

The accuracy of the invasive margin annotations performed by the researcher performing the image analysis (CB) were validated via statistical comparison to annotations made by a board – certified, specialist gastrointestinal pathologist blinded to the image analysis study who had performed KM grading in previous studies (Karin Oien). 35 sections from the TransSCOT clinical trial cohort were chosen at random and annotated for the invasive margin by the pathologist, with only the aim of the study and a brief description of how the invasive margin annotations had been performed by the image analysis researcher for other studies. The high threshold lymphocyte detection algorithm was then applied to both researchers' annotations and patients stratified for high vs low immune infiltrate using the threshold determined in the Glasgow Development cohort (Figure 4.4). Cohen's Kappa metric was used to determine the level of statistical agreement between the two researchers, producing a value of $\kappa = 0.81$, indicating excellent concordance (Table 4.2).

		GI Pathologist (KO)		
		High Immune Infiltrate	*Low Immune Infiltrate*	***Cohen's Kappa***
Image Analysis Researcher (CB)	*High Immune Infiltrate*	22	2	**0.81**
	Low Immune Infiltrate	1	10	

Table 4.2 Statistical Agreement of Invasive Margin Annotations Between Researchers. *Confusion matrix and Cohen's Kappa score of agreement between invasive margin annotations stratified by image analysis lymphocyte density.*

4.3.2 Validation of Image Processing Lymphocyte Detection Algorithm to the Glasgow Validation Cohort

Given the superior performance of the High Threshold lymphocyte detection algorithm in cell segmentation accuracy, correlation with manual histopathological assessment, and association with CSS compared to the Initial and Low threshold algorithms, this algorithm was selected to be validated in further CRC patient cohorts, initially the Glasgow Validation Cohort. In this cohort, multiple sections were available in varying numbers for each patient (N = 1 – 6), but it was unknown which section from each patient was originally used for histopathological assessment. Therefore, the image analysis researcher reviewed all available sections for each patient and determined the section that showed the deepest point of tumour invasion as per the KM grading criteria (Klintrup et al., 2005). However, as the sections had been marked up by a pathologist for teaching and TMA construction, the scanning quality of the sections was lacking due to focussing issues arising from the pen markings on the coverslips, meaning an additional consideration had to be made as to how appropriate the section was for image analysis. It was decided that if the section showing the deepest point of tumour invasion was not of appropriate quality for image analysis, another slide was chosen for that patient contingent on the slide being of the same T – Stage as the slide showing the deepest point of invasion, and if no slides met these criteria, the patient was excluded from analysis. This resulted in 521 patients being included in this study.

The selected sections were annotated for the invasive margin by the image analysis researcher to generate analysis ROI and the high threshold lymphocyte detection algorithm detailed in 4.3.1 was applied within these ROI. The cell densities determined by the algorithm for each patient were compared to manual histopathological KM grading by ROC analysis. Plotting the algorithm determined cell densities against the KM grade stratification produced an AUC = 0.702 (Figure 4.6).

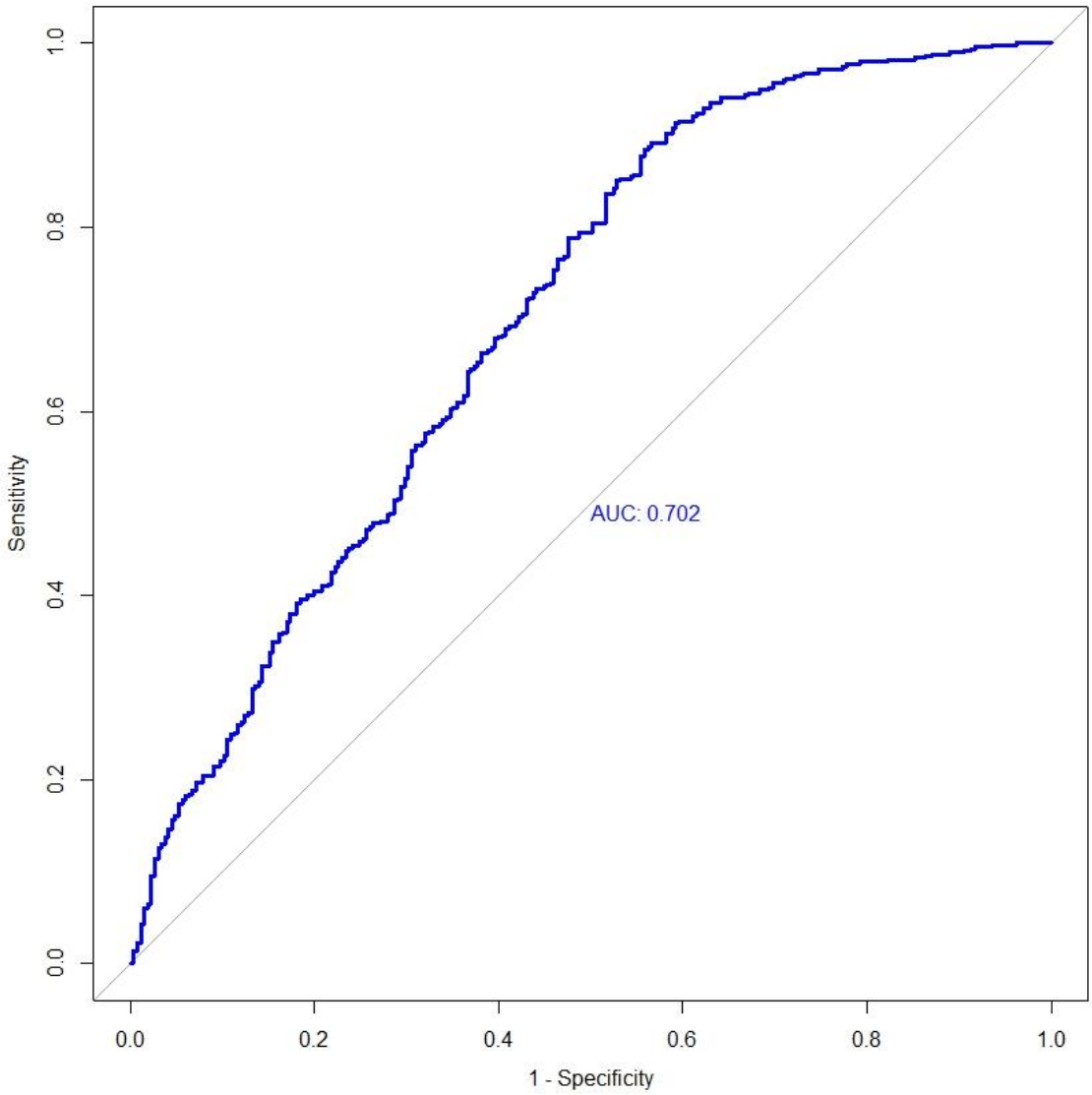


Figure 4.6 Receiver Operator Characteristic of Lymphocyte Density in Glasgow Validation Cohort. *ROC curve for lymphocyte detection algorithm determined cell density against Klintrup – Mäkinen grading stratification in the Glasgow Development Cohort, AUC = 0.702.*

The cell densities determined by the lymphocyte detection algorithm were used to stratify patients into high vs low immune infiltrate based on the same threshold determined in the Glasgow Development cohort, 0.88 cells / mm^2 (Figure 4.4). Kaplan – Meier survival analysis of the stratified cell densities showed a statistically significant correlation with CSS (HR 1.499 95% CI 1.034 – 2.175 *P = 0.032*, Figure 4.7).

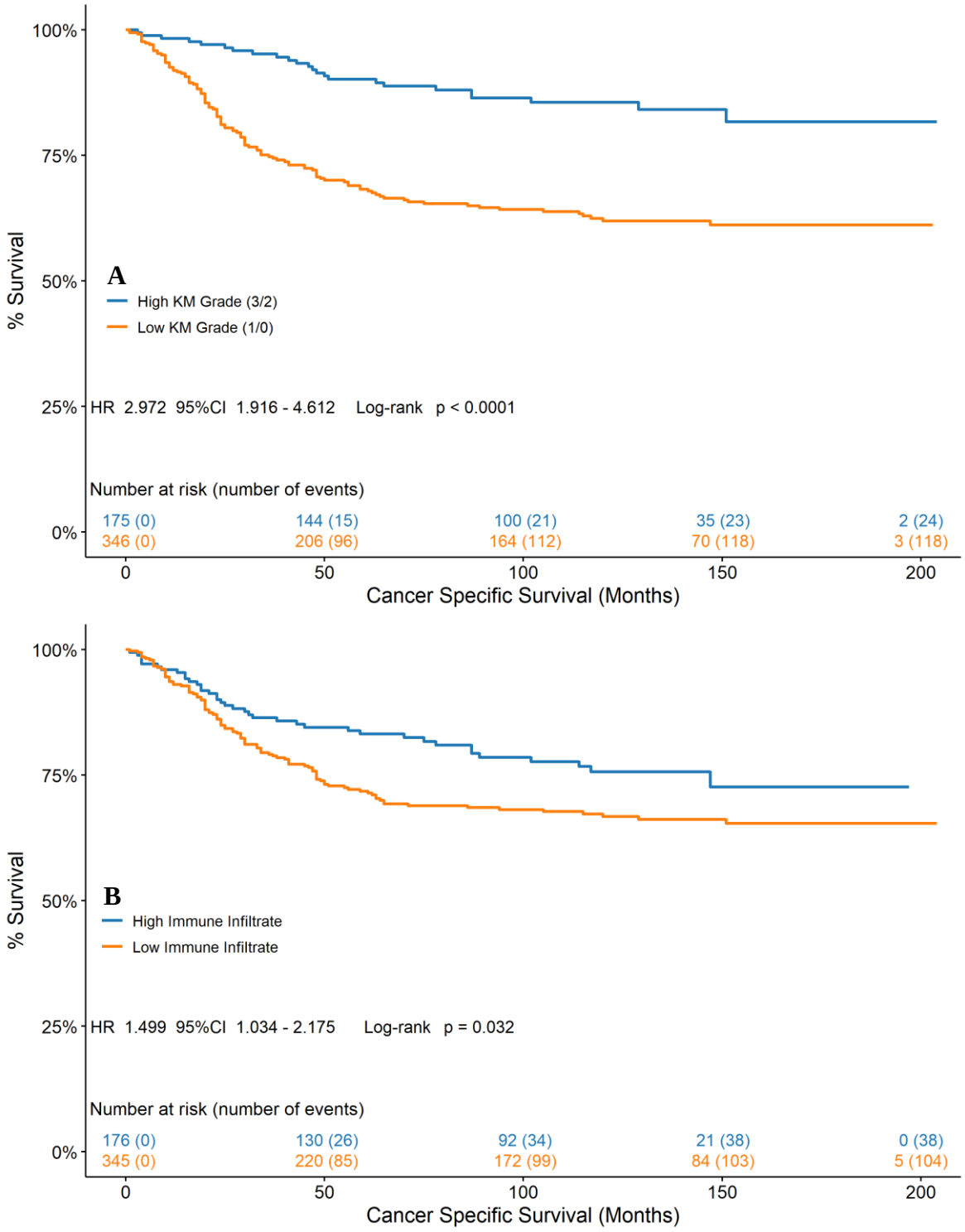


Figure 4.7 Relationship between CSS and Lymphocyte Density in Glasgow Vadliation Cohort. *Association of CSS and lymphocyte density determined by histopathological Klintrup – Mäkinen grading (A) and image analysis lymphocyte detection (B) in the Glasgow Validation Cohort.*

4.3.3 Assessment of Peritumoural Lymphocyte Response via Image Analysis in TransSCOT Clinical Trial Cohort

The lymphocyte detection algorithm was developed and initially validated in two geographical and process similar but chronologically independent cohorts, the Glasgow Development and Validation cohorts. Given the thresholding methods used to develop the algorithm, it is inherently susceptible to stochastic changes in stain vectors arising from a multitude of processing variations that can occur during section production, such as the use of different types of haematoxylin, FFPE block age and storage, and the slide scanning process. Therefore, the algorithm was further validated in a larger, geographically and process independent cohort, the TransSCOT Clinical Trial cohort (detailed in 2.1.6). The image analysis method employed was identical to that of the previous cohorts; the image analysis researcher (CB) annotated all H&E sections in the cohort for the invasive margin to produce an ROI for analysis, within which the high threshold lymphocyte detector was applied. The lymphocyte density for each patient was calculated and ROC analysis was used to assess the agreement with manual histopathological KM grading. The lymphocyte densities were plotted against the stratified KM grades and produced and AUC = 0.699 (Figure 4.8).

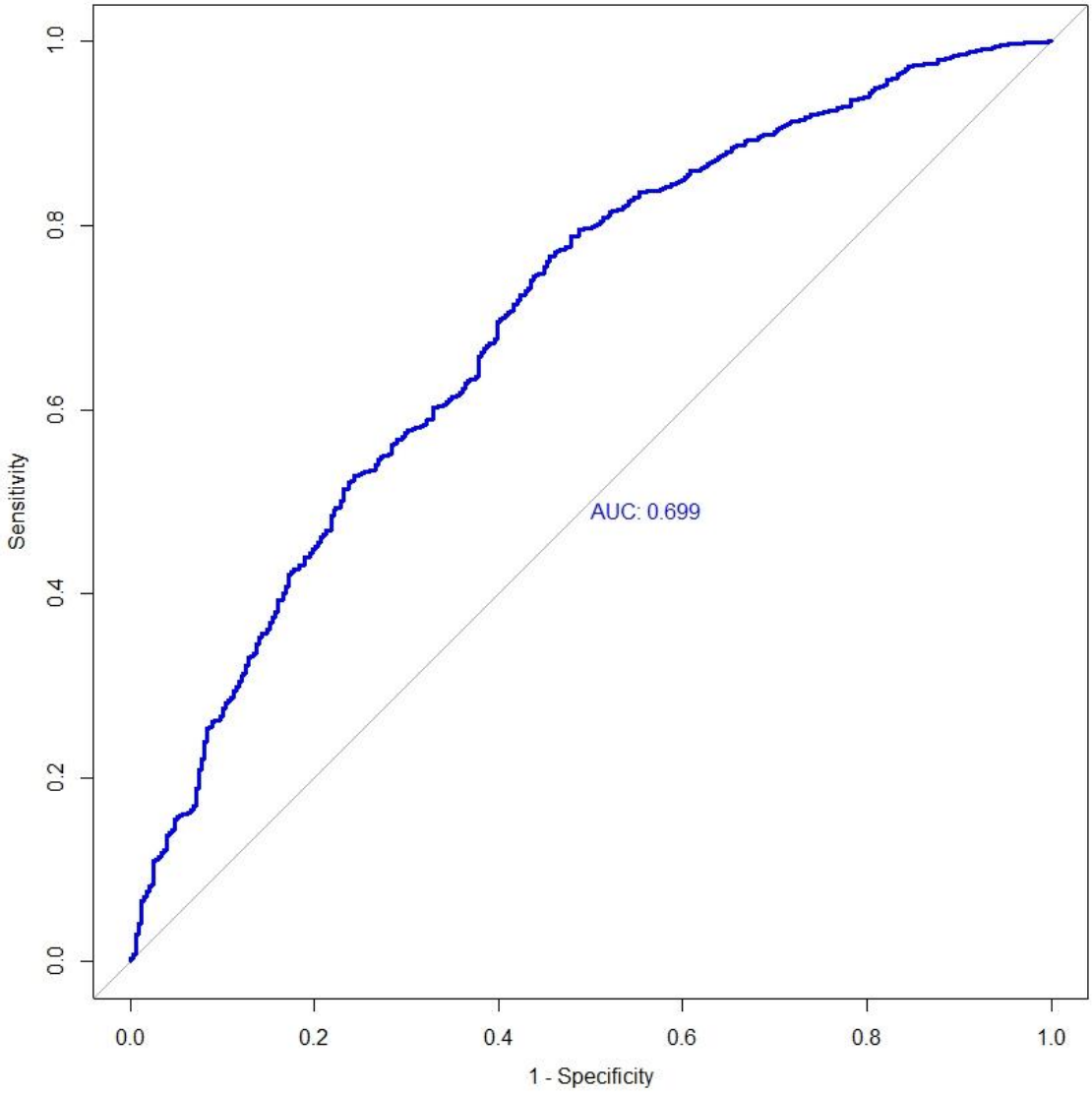


Figure 4.8 Receiver Operator Characteristic of Lymphocyte Density in TransSCOT Clinical Trial Cohort. *ROC curve of image analysis determined lymphocyte densities against KM grade stratification in TransSCOT Clinical Trial cohort.*

The lymphocyte densities determined by the algorithm were used to stratify patients into high vs low immune infiltrate based on the same cut off value determined in the Glasgow Development cohort (0.88 cells / mm^2, Figure 4.4). Kaplan – Meier survival analysis was used to determine the association of the stratified lymphocyte densities with DFS and to compare this association to that of histopathological KM grading. The algorithm determined lymphocyte densities showed a statistically significant association with DFS, performing comparably to manual KM grading (HR 1.46 95% CI 1.14 – 1.87 *P = 0.0025*, Figure 4.9).

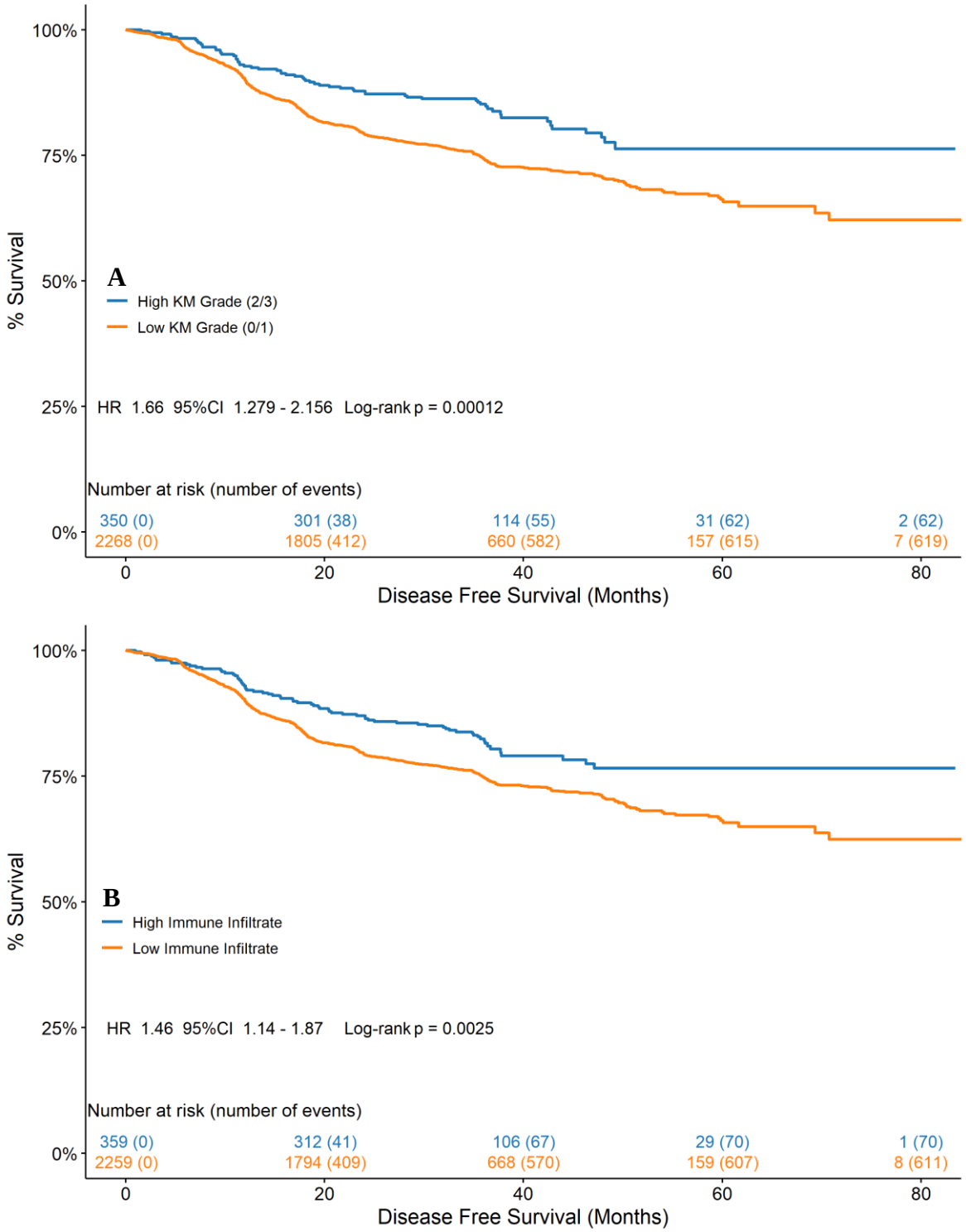


Figure 4.9 Relationship Between DFS and Lymphocyte Density in TransSCOT Clinical Trial Cohort. *Association between DFS and lymphocyte density assessed via Klintrup – Mäkinen grading (A) and Image Analysis (B) in the TransSCOT Clinical Trial cohort.*

Univariate and multivariate cox regression was performed to assess the relationship between the peritumoural inflammatory response determined via image analysis, clinicopathological features, and DFS (Table 4.3). On multivariate analysis, the peritumoural lymphocyte density was found to be significantly associated with DFS (HR 0.731 95%CI 0.579-0.929, $p = 0.01$), independent of T-Stage ($p < 0.001$) and N-Stage ($p < 0.001$).

Univariate				**Multivariate**			
	HR	95% CI	p-value		HR	95% CI	p-value
Sex							
Female	1.0						
Male	1.040	0.898-1.205	0.599				
T Stage				**T Stage**			
1	1.0			1	1.0		
2	1.013	0.498-2.060	0.973	2	3.799	0.906-15.927	0.068
3	1.85	0.988-3.467	0.055	3	6.530	1.626-26.229	**0.008**
4	3.217	1.715-6.037	**<0.001**	4	11.414	2.834 -45.960	**<0.001**
N Stage				**N Stage**			
0	1.0			0	1.0		
1	1.436	1.155-1.787	**0.001**	1	1.858	1.472-2.346	**<0.001**
2	2.821	2.249-3.538	**<0.001**	2	3.081	2.423-3.918	**<0.001**
Site				**Site**			
Left	1.0			Left	1.0		
Right	0.687	0.557-0.848	**<0.001**	Right	0.827	0.659-1.038	0.101
Treatment							
FOLFOX	1.0						
CAPOX	1.084	0.924-1.272	0.319				
Image Analysis Lymphocyte Density				**Image Analysis Lymphocyte Density**			
Low	1.0			Low	1.0		
High	0.669	0.527-0.849	**<0.001**	High	0.731	0.576-0.929	**0.01**

Table 4.4 Relationship between Peritumoural Lymphocyte Densty, Clinicopathological Features and DFS in TransSCOT Cohort. Hazard ratios and 95% confidence intervals determined via Cox proportional hazards regression and multivariate analysis conducted using backwards conditional method.

4.3.4 Development of a Fully Automated Image Analysis Approach to Tumour Infiltrating Lymphocyte Assessment

The lymphocyte assessment workflow detailed and validated in 4.3.1 – 4.3.3 demonstrated that image analysis detection of lymphocytes on clinical H&E sections can produce reliable, prognostically significant patient stratification. However, the workflow requires manual annotation of the invasive margin, demanding time from the researcher / pathologist and introducing a potential source of bias due to the subjective nature of invasive margin assessment, which would require additional time and a second observer to account for. Attempting to reduce subjectivity within the assessment and reduce the time requirement of the analysis, the aim of the present section is to develop a fully automated workflow for lymphocyte assessment on H&E sections utilising the deep learning TSP algorithm detailed in 3.3.3. This assessment would therefore be conducted across the whole of the tumour area within the tumour associated stroma instead of at the invasive margin, as it was not viable to reliably automate detection of the invasive margin without meticulous curation of the sections utilised in the analysis which would ultimately reduce the power of the study beyond an acceptable degree. While it is understood that lymphocyte infiltrate assessment within the tumour core can produce prognostically significant results when conducted through IHC based methods, H&E based assessment holds significantly more prognostic value when conducted at the invasive margin over the central tumour (Alexander et al., 2020). Therefore, it is of interest as to whether a quantitative image analysis approach to TIL assessment on CRC H&E sections could be used to reliably stratify patients for survival.

Initially, the deep learning – based TSP workflow detailed in Figure 3.12 was applied to the H&E section, beginning with the tissue isolation algorithm (2.4.3.1) to remove processing artefacts and slide glass background, followed by the full TSP algorithm. The areas quantified as stroma by the algorithm were subsequently converted into ROI within which the high threshold lymphocyte detection algorithm was applied, and the cell densities in cells / mm^2 were determined. The workflow for this algorithm is detailed in Figure 4.10.

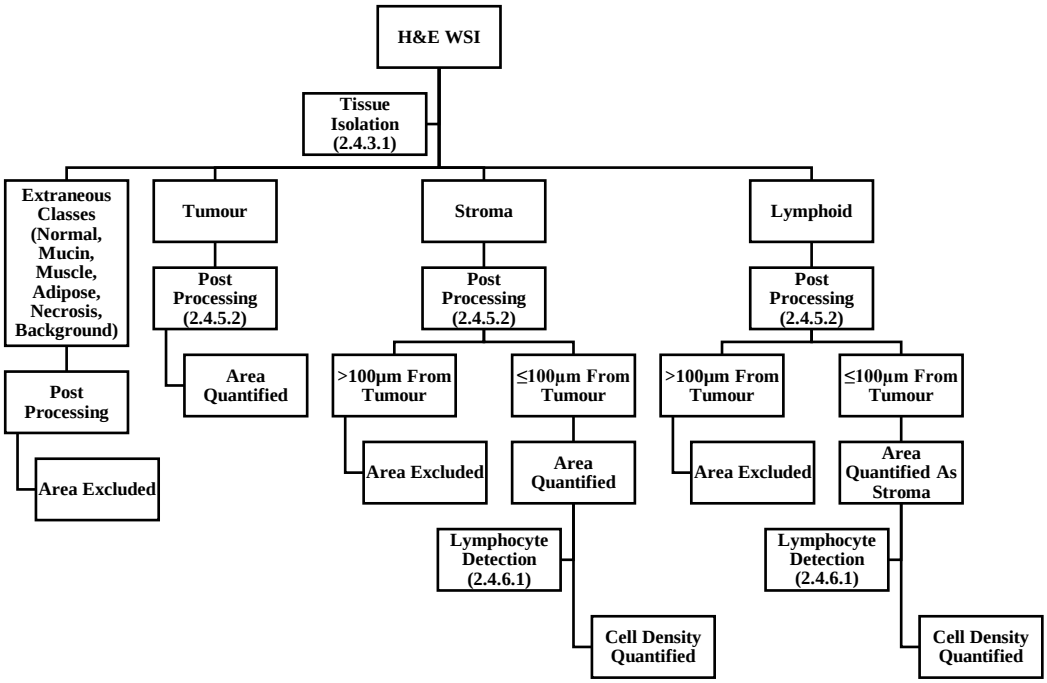


Figure 4.10 Automated Lymphocyte Detection Algorithm Workflow. *Workflow of automated lymphocyte density image analysis algorithm coupling deep learning TSP and TIL assessment. Methods sections detailing operations are given in brackets.*

To validate the TSP coupled lymphocyte detection algorithm for prognostic significance, the full workflow was applied to all H&E sections from the Glasgow Development cohort. CD3 immunohistochemistry had been previously performed and scored in this cohort and patient stratification for high vs low $CD3^+$ cell infiltrate based on the median expression value was available in the database, therefore, the stromal TIL densities determined by the algorithm for each patient underwent ROC analysis against the median CD3 stratification to assess the association. The ROC analysis between H&E stromal TILs and CD3 IHC produced and AUC = 0.724 (Figure 4.11).

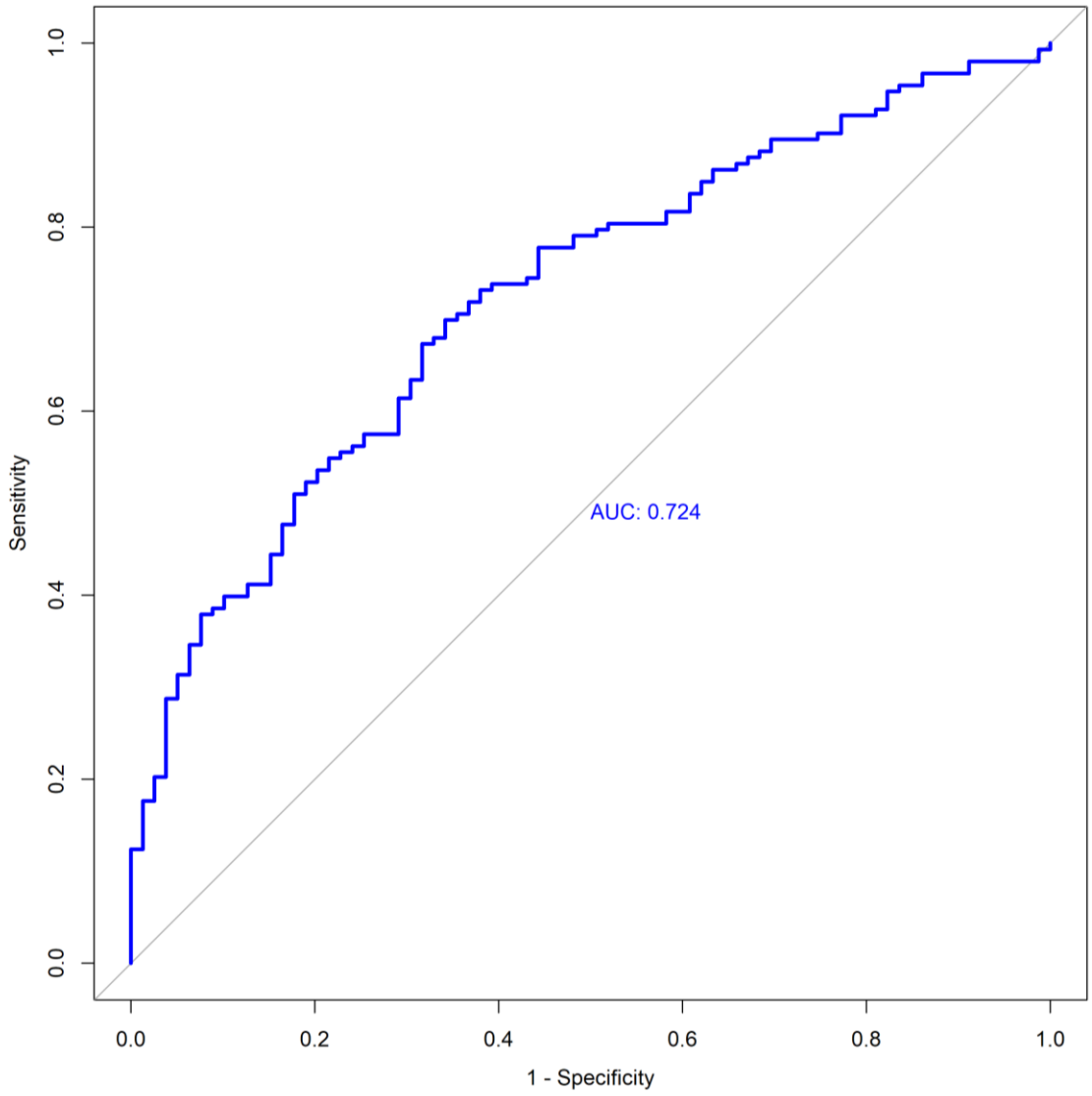


Figure 4.11 Receiver Operator Characteristic of Stromal TILs in Glasgow Development Cohort. *ROC curve of stromal TIL density determined via fully automated image analysis on H&E sections against median stromal CD3$^+$ cell count in the Glasgow Development Cohort.*

To compare H&E stromal TILs density to CD3 IHC for prognostic significance, an optimal cut off value based on CSS was determined using the survminer and maxstat packages in R. The cut off value determined for this algorithm and cohort was 0.55 cells / mm^2 (Figure 4.12).

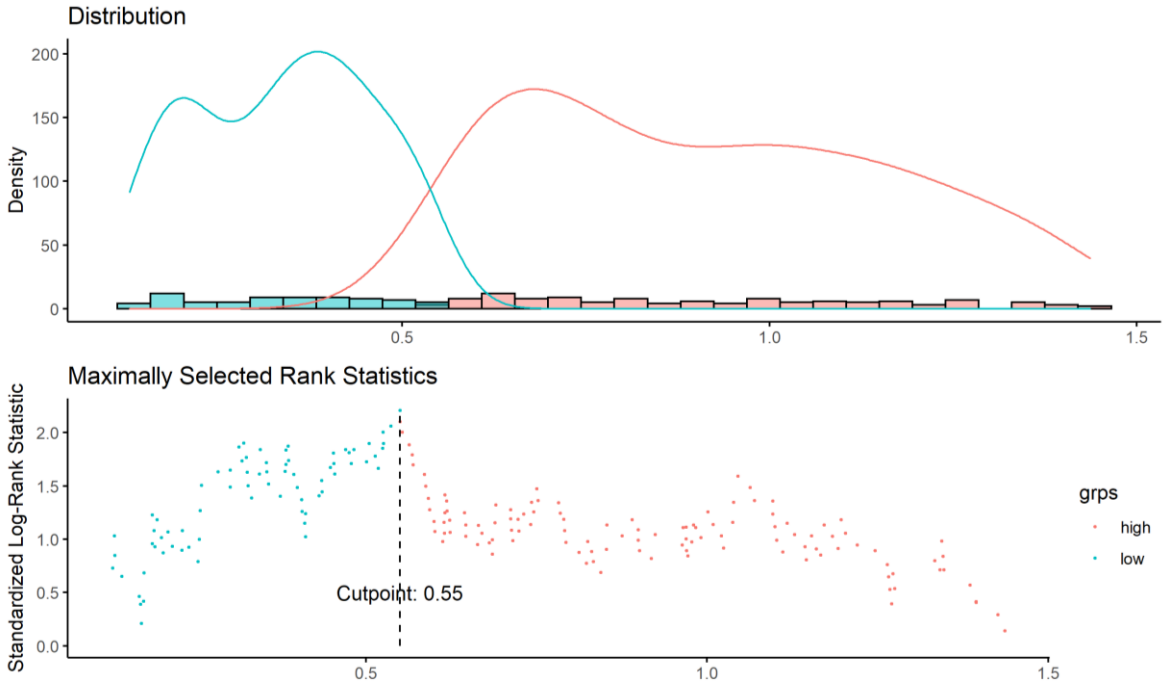


Figure 4.12 Determining Optimal Cut Off Value For Stromal TILs in Glasgow Development Cohort. *Distribution and optimal cut off value stromal TIL density determined via automated image analysis for cancer specific survival in Glasgow Development Cohort.*

The cut off value (0.55 cells / mm^2) was used to stratify patients into high vs low immune infiltrate. Kaplan – Meier survival analysis was used to assess the association of the stratified TIL densities with CSS and compare this to the stratified CD3 IHC scores. The image analysis TIL density assessment showed a statistically significant association with CSS in this cohort (HR 1.637 95%CI 1.056 – 2.538 *P = 0.026*, Figure 4.13).

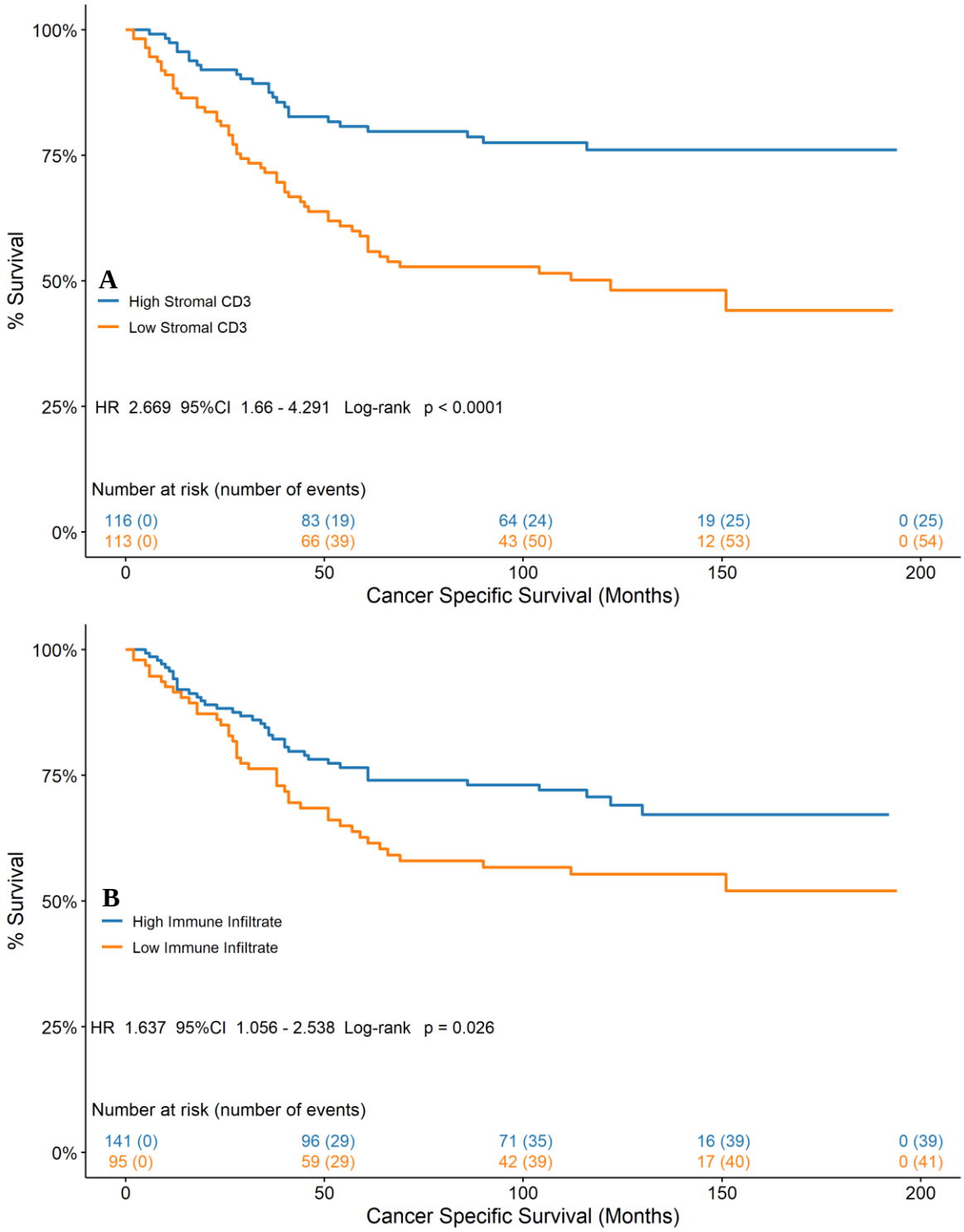


Figure 4.13 Relationship Between Tumour Infiltrating Lymphocyte Density and CSS in the Glasgow Development Cohort. *Association between CSS and TIL density assessed via CD3 IHC (A) and automated H&E based image analysis in the Glasgow Development Cohort.*

4.3.5 Validation of Automated Image Analysis TIL Assessment in the Glasgow Validation Cohort

To assess the performance of the automated lymphocyte detection algorithm in a larger cohort, the workflow was subsequently validated in the Glasgow Validation cohort. The algorithm was applied to the sections determined during validation of the annotated invasive margin analysis detailed in 4.3.2. CD3 immunohistochemistry had also been performed on full sections in this cohort and the median score for $CD3^+$ cells was used to stratify patients for CSS. The cell densities determined for each patient by the automated algorithm underwent ROC analysis against the median stratified CD3 scores and generated and AUC = 0.688 (Figure 4.14).

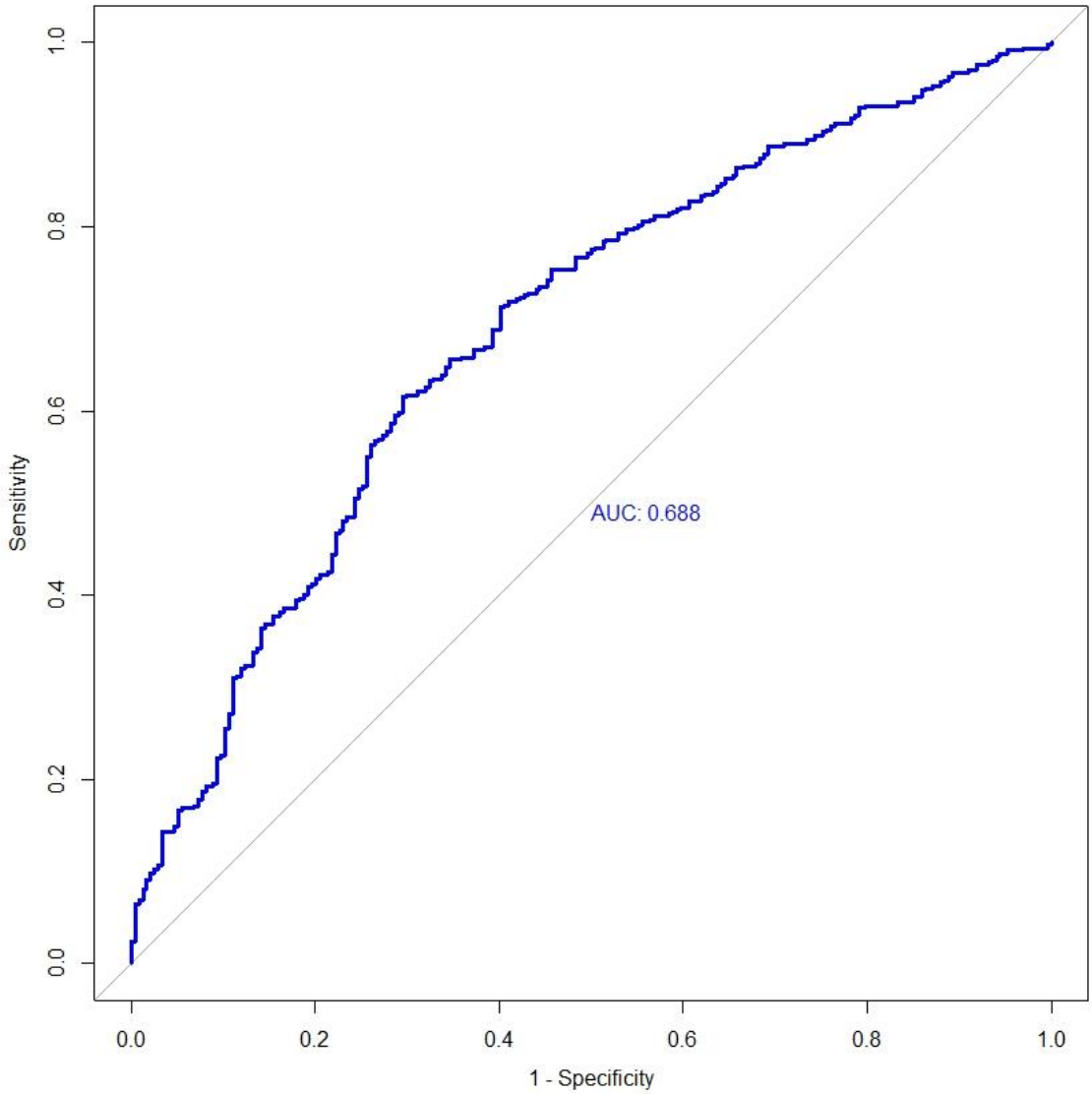


Figure 4.14 Receiver Operator Characteristic of TIL Density in Glasgow Validation Cohort. *ROC curve of automated image analysis determined TIL densities against median stratified CD3 scores in the Glasgow Validation Cohort.*

The automated algorithm determined TIL densities were used to stratify patients for high vs low immune infiltrate based on the same cut off value determined in the Glasgow Development cohort in 4.3.4, 0.55 cells / mm^2. Kaplan – Meier survival analysis was used to assess the relationship between the stratified TIL densities and CSS and to compare this association to that of manually assessed CD3 densities. The image analysis determined TIL densities showed a non – statistically significant association with CSS but with a hazard ratio of 1.183, suggesting an improved prognosis for high immune infiltrate patients. The automated image analysis did, however, show a marginally better association with CSS than histopathological CD3 assessment (HR 1.183 95%CI 0.8178 – 1.711 *P = 0.37* vs HR 1.135 95% 0.7967 – 1.616 *P = 0.48*, respectively, Figure 4.15).

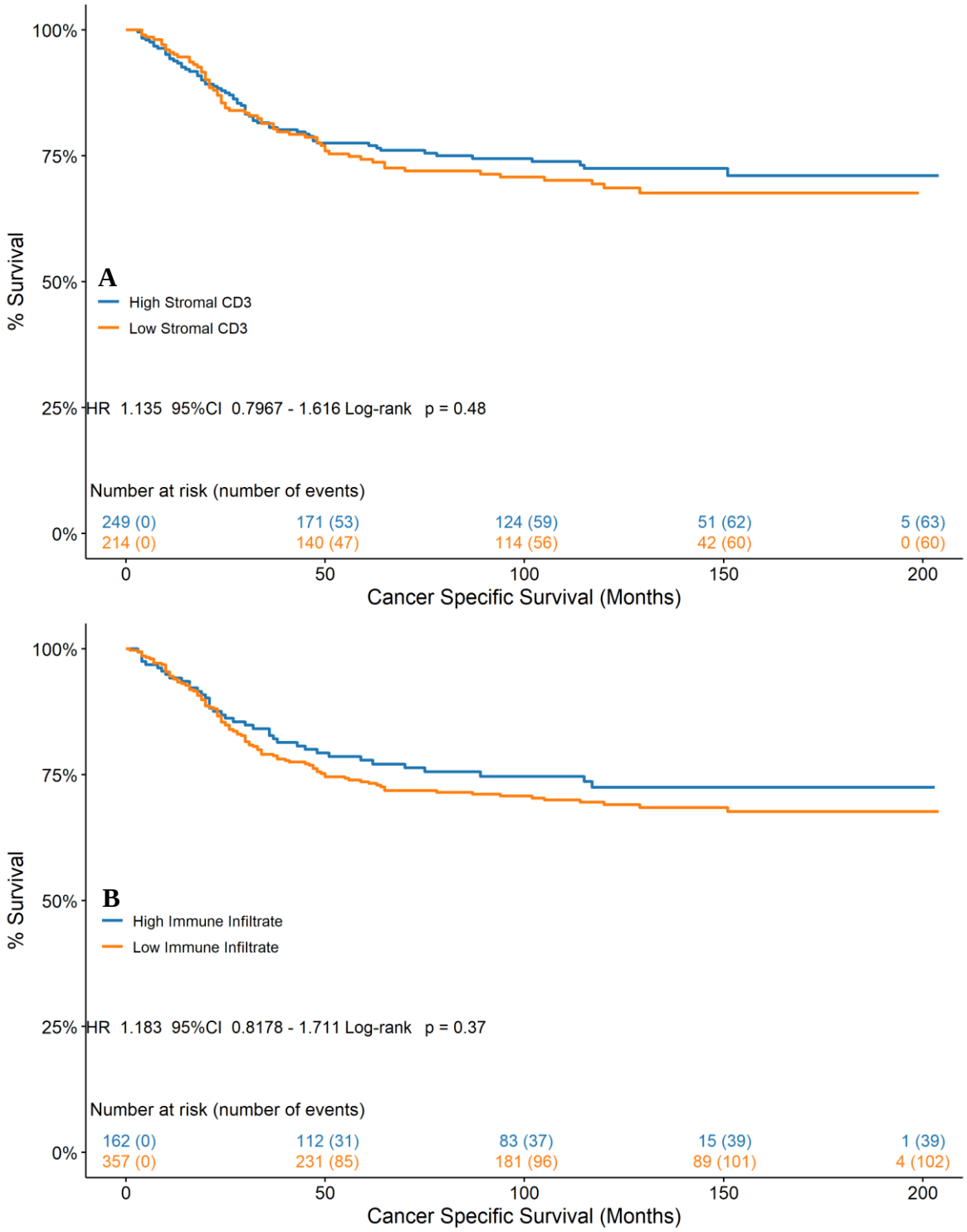


Figure 4.15 Relationship Between Tumour Infiltrating Lymphocyte Density and CSS in the Glasgow Validation Cohort. *Association between CSS and TIL density assessed via CD3 IHC (A) and automated H&E based image analysis in the Glasgow Validation Cohort.*

4.3.6 Assessment of TIL Densities via Automated Image Analysis in a Previously Unscored Patient Cohort

Clinical translation of image analysis algorithms is dependent on their ability to maintain classification performance and prognostic significance when presented with novel image data from new patient cohorts. While the TransSCOT Clinical Trial cohort has been previously assessed for the TSP and KM grade via manual histopathological assessment and image analysis using the components of the automated lymphocyte detection workflow (3.3.5 and 4.3.3 respectively), no previous assessment of the stromal TIL density had been conducted, on H&E sections or through IHC methods.

To validate the automated TIL workflow in this cohort and to determine the prognostic effect of TILs in a clinical trial setting, the full lymphocyte detection workflow was applied to a single H&E section from each patient, totalling 2856 after thirty – day mortality exclusion. The cell densities determined by the algorithm for each patient were stratified into high vs low immune infiltrate using the same cut off value determined in the Glasgow Development Cohort, 0.55 cells / mm^2. Kaplan – Meier survival analysis showed that the stratified TIL densities had statistically significant association with DFS in this cohort (HR 1.34 95%CI 1.159 – 1.549 $P < 0.0001$, Figure 4.16).

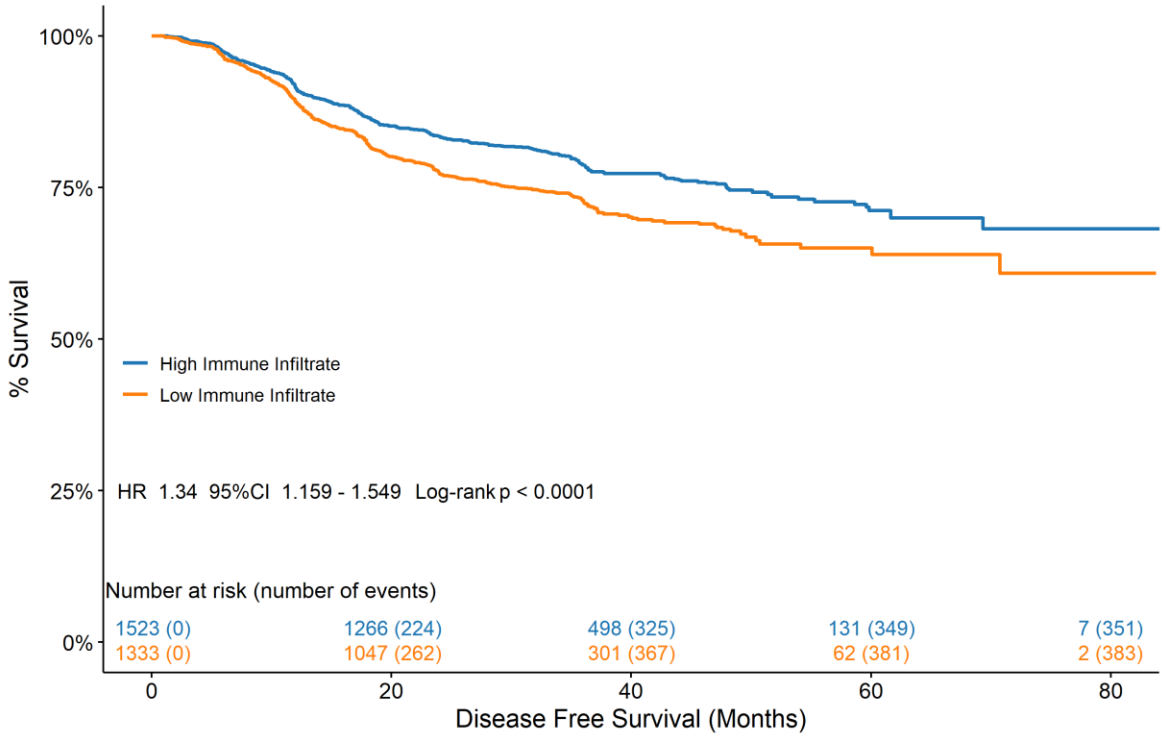


Figure 4.16 Relationship Between TIL Density and DFS in TransSCOT Cohort. *Association of DFS and TIL density determined through automated image analysis in the TransSCOT Clinical Trial Cohort.*

Univariate and multivariate cox regression was performed to assess the relationship between TILs determined via image analysis, clinicopathological features, and DFS (Table 4.5). On multivariate analysis, TILs were found to be significantly associated with DFS (HR 0.804 95%CI 0.696-0.994, $p = 0.003$), independent of T-Stage ($p < 0.001$) and N-Stage ($p < 0.001$).

Univariate				Multivariate			
	HR	95% CI	p-value		HR	95% CI	p-value
Sex							
Female	1.0						
Male	1.040	0.898-1.205	0.599				
T Stage				**T Stage**			
1	1.0			1	1.0		
2	1.013	0.498-2.060	0.973	2	1.089	0.535-2.215	0.815
3	1.85	0.988-3.467	0.055	3	1.867	0.995-3.504	0.052
4	3.217	1.715-6.037	**<0.001**	4	3.364	1.784-6.342	**<0.001**
N Stage				**N Stage**			
0	1.0			0	1.0		
1	1.436	1.155-1.787	**0.001**	1	1.825	1.458-2.285	**<0.001**
2	2.821	2.249-3.538	**<0.001**	2	3.159	2.511-3.974	**<0.001**
Site				**Site**			
Left	1.0			Left	1.0		
Right	0.687	0.557-0.848	**<0.001**	Right	0.800	0.644-0.994	0.044
Treatment							
FOLFOX	1.0						
CAPOX	1.084	0.924-1.272	0.319				
TILs				**TILs**			
Low	1.0			Low	1.0		
High	0.779	0.675-0.900	**<0.001**	High	0.804	0.696-0.928	**0.003**

Table 4.5 Relationship between TILs, Clinicopathological Features and DFS in TransSCOT Cohort. Hazard ratios and 95% confidence intervals determined via Cox proportional hazards regression and multivariate analysis conducted using backwards conditional method.

4.4 Discussion

The local inflammatory response to CRC has long been understood to hold significant prognostic value. The effect of a conspicuous peritumoural immune cell infiltrate on patient survival was first described and characterised in 1931 in rectal tumours (Maccarty, 1931a), however, little progress was made in further investigating the utility of this assessment until later in the 20th century. Though some studies sought to establish the host immune response to CRC as a meaningful criterion for diagnosis, it was not until the work of J.R. Jass that reproducible means of assessing the infiltration of immune cells would be developed (C. S. D. Roxburgh & McMillan, 2012). Jass semi quantitatively assessed lymphocytic infiltration of 447 H&E sections of rectal tumours, showing that a marked difference in peritumoural lymphocyte densities conferred a 5 – year survival difference of 92% for patients with a pronounced response, and 36% for patients with a weak response independent of disease stage (Jass, 1986). Despite showing good reproducibility ($\kappa = 0.72$), being subsequently validated for prognosis in an independent cohort, and demonstrating further prognostic utility as part of a novel classification system, Jass' criteria saw no attempts at incorporation into routine pathology (Jass et al., 1987; C. S. D. Roxburgh et al., 2009a, 2009b).

Continuing research into defining how the local inflammatory response impacts cancer progression led subsequent studies to develop yet more reproducible methodologies for tissue – based immune assessment. The Immunoscore, developed by Galon and Colleagues, showed excellent prognostic utility by subtyping lymphocyte populations through IHC and quantification through digital pathology, however, the costly implementation of diagnostic IHC for each patient coupled with proprietary image analysis software prohibited clinical use of the system (Galon et al., 2006, 2014). The most clinically translatable method of lymphocyte assessment is the criteria developed by Klintrup, Mäkinen and Colleagues, whereby lymphocytic infiltrate is semi quantitatively assessed on H&E sections according to a four – point grading system (Klintrup et al., 2005). The KM grading system as it became known, is highly prognostic for survival, is reproducible between both observers and institutions, has no additional associated cost, and has been validated in a number of independent studies (Park et al., 2017; Park, McMillan, et al., 2016; Richards et al., 2012, 2014; C. S. D. Roxburgh et al., 2009b), however, it too has seen no clinical implementation to date with the lack of a robust, comprehensive, and standardised assessment criteria cited as the primary reason. It is of interest therefore, to investigate the utility of image analysis to conduct these assessments as a single observer with the aim of ensuring their reliability.

Image analysis has been utilised in prior studies to quantify immune cell densities and determine their prognostic effect; however, this has primarily been conducted on IHC labelled sections. While conducting reliable image analysis on H&E sections is difficult due to the histological variability between labs, the recent push towards histological standardization associated with routine diagnostic reporting from WSI has resulted in an ever – increasing number of slides of sufficiently consistent quality to make H&E WSI an appropriate modality for image analysis. Therefore, to somewhat retrospectively assess the feasibility of translating the KM criteria to an image analysis – based assessment, the present study assessed the prognostic capacity of image analysis utilising H&E sections produced during routine diagnosis.

Given that a training data set of pathologist annotated lymphocytes was not available to train an artificial intelligence algorithm, this study was conducted using traditional, threshold – based image analysis on operator crafted features to detect lymphocytes. An initial algorithm was produced using thresholds determined by the image analysis researcher to isolate lymphocytes based on visual assessment of the algorithm's performance across a series of histologically variable H&E sections until satisfactory lymphocyte detection was achieved. This approach was validated using two methodologies; DSC analysis against IHC sections stained for CD3 and for patient prognosis against manual KM assessment.

The initial algorithm was comparatively validated against two additional algorithms, infra and supra optimisations; one where the detection thresholds were reduced resulting in more cell types and numbers to be detected, and one where the same thresholds were increased resulting in fewer cells being detected. The theory behind this being that detecting more cells of more types would reduce the prognostic significance of the algorithm and could either increase or decrease the DSC values depending on the ratio of additional cells detected and increasing the thresholds would produce an identical or better prognosis and again, produce an unknown effect on the DSC values. Indeed, in patients from the Glasgow Development cohort, increasing the threshold values to reduce the number of detected cells produced the best DSC value against the initial and low threshold algorithms (Figure 4.2), the greatest association with CSS (Figure 4.5), and the best statistical agreement with manual KM assessment (Figure 4.3, Table 4.1), and was therefore chosen as the optimal algorithm to conduct assessment in subsequent cohorts. Prior to applying the lymphocyte detection workflow to subsequent cohorts, the accuracy of the researcher's invasive margin annotations made to validate the lymphocyte detection algorithm was determined against those of a specialist pathologist. Ideally, DSC values would have been generated to directly

compare the annotations, however this was not possible, therefore, the high threshold algorithm (hereafter referred to simply as the lymphocyte detection algorithm) was applied within both observers IM annotations and the kappa metric for the stratified lymphocyte densities showed excellent concordance between the two observers, $\kappa = 0.81$. It cannot be definitively stated that this was due to the annotations being in identical locations on the section, however, if the annotations were in differing sites this did not appear to have a significant effect on the analysis, suggesting that annotation of a prognostically appropriate area by an experienced observer is equally sufficient for conducting this analysis.

While the lymphocyte detection algorithm coupled with invasive margin annotation demonstrated prognostic significance in the Glasgow Development cohort in which it was developed (Figure 4.5 D), though to a lesser degree than manual histopathological assessment (Figure 4.5 A), traditional threshold – based image analysis algorithms utilising colour – derived features do not translate well between independent cohorts. Data from the Glasgow Validation cohort demonstrated the lymphocyte detection algorithm still maintained a statistically significant association with CSS, (Figure 4.7), although again to a notably lesser degree than manual histopathological assessment (Figure 4.7). This could potentially be due to issues discussed in 3.4 regarding the quality of sections from this cohort, as data from the TransSCOT clinical trial cohort, the largest cohort and the one with the most recently produced WSI, demonstrated that the lymphocyte detection algorithm performed comparably to manual histopathological assessment in terms of patient stratification for DFS (Figure 4.7).

These data demonstrate the feasibility of semi – automated image analysis for conducting peritumoural lymphocytic infiltrate assessment to prognostically stratify patients. However, it is noteworthy that in each instance that the image analysis underperforms for CSS stratification compared to manual histopathological assessment, though the underlying reasons for this are not immediately apparent, the most likely explanation is in the translation of a qualitative morphology – based assessment to a quantitative density – based assessment. Klintrup, Mäkinen and Colleagues conducted an exhaustive study of the patterns of peritumoural lymphocyte response and the relation to prognosis, however, the criteria describing the patterns of a “patchy”, “band – like”, and “florid cup” response do not directly translate to cell density analysis, which serves to act only as a surrogate for these measures. It is highly likely that the profound prognostic significance of the KM grading system lies within the nuances of these criteria which is not recapitulated through the current image analysis approach, hence why image analysis comparatively underperforms for survival

stratification. Recent studies have identified the prognostic significance of the spatial relationships between the tumour and its microenvironment, particularly infiltrating lymphocytes, which could provide a means by which to more directly translate the KM grading criteria to an image analysis – based assessment (Corredor et al., 2019; I. Nearchou et al., 2021; I. P. Nearchou, Lillard, et al., 2019b).

The semi – automated lymphocyte detection workflow demonstrates good prognostic significance for survival but only partially addresses the requirements of a clinically translatable image analysis algorithm. While the cell detection and quantification are not subject to inter observer variability, there is still a time requirement and level of subjectivity associated with the analysis in the annotation of the invasive margin. Therefore, a fully automated approach to lymphocyte density analysis utilising the TSP algorithm developed in Chapter 3 was applied to the same sections as the semi – automated approach to assess the level to which it addressed these issues. Data from the Glasgow development cohort demonstrated that the fully automated approach to TIL assessment significantly stratified patients for CSS (Figure 4.13 B), though to a lesser degree than that of CD3 IHC manual histopathological assessment (Figure 4.13 A). This is most likely because CD3 IHC isolates T lymphocytes which are known to have the greatest effect on prognosis, whereas the H&E – based image analysis detects all lymphocytes thus diluting the prognostic effect. In the Glasgow Validation cohort however, the automated image analysis outperformed the CD3 IHC assessment for CSS stratification though neither reached statistical significance (Figure 4.15). Given the data from the Glasgow Development Cohort, it would be expected that the CD3 IHC significantly stratified patients to a similar degree in the Glasgow Validation cohort, the fact that it did not is likely due to the use of the median $CD3^+$ cell density value to dichotomize patients as opposed to a cut off value optimised for CSS. The prognostic ability of automated lymphocyte analysis was shown however in the TransSCOT clinical trial cohort, which had not been assessed for CD3, where the algorithm demonstrated a significant association with DFS (Figure 4.16).

While the automated assessment showed associations with survival, it failed to reach the same significance as the semi – automated assessment, most likely due to the fact that H&E based assessment of lymphocytic infiltrate is most prognostic at the invasive margin and assessment within the tumour core requires IHC subtyping of lymphocytes to elucidate the effect individual populations have on prognosis (Alexander et al., 2020). Deep learning could provide a solution to the limitations of the present study, in the case of the semi – automated assessment a CNN could be trained to not only identify lymphocytes from H&E

sections but also delineate specific immune cell subtypes to isolate their prognostic effect. Additionally, automated detection of the invasive margin to further reduce the subjectivity of the assessment could be achievable without deep learning but would require significant curation of the WSIs utilised to be applicable across multiple patient cohorts, therefore, deep learning would be a logical solution to both issues.

In summary, these data demonstrate the feasibility of using multiple image analysis methodologies to assess lymphocytic infiltrate from clinical H&E sections and their use in predicting survival outcomes across multiple cohorts. Although requiring additional time to conduct, semi – automated assessment of the peritumoural immune response yields prognostic stratification comparable to that of manual histopathological assessment but will need further investigation to determine the specific prognostic relationships described by the KM criteria. Fully automated assessment requires no input from the pathologist but reduces the prognostic ability of the assessment. Thus, a determination of the trade – off between time and quality of stratification is necessary when deciding which methodology should be implemented in a given study.

5. Automated Assessment of Ki67 Expression via Image Analysis in CRC

5.1 Introduction

Immunohistochemistry is a quick and relatively inexpensive technique routinely used in diagnostic pathology to provide additional information to clinicians about the nature of the specific malignancy. In colorectal cancer, IHC is regularly performed to assess the mismatch repair (MMR) status of the patient by staining for four proteins known to be involved in the MMR process (Richman, 2015). While IHC has been implemented in routine pathology across the world, there are many prognostic and predictive biomarkers that are readily assessable through IHC that have yet to be incorporated into diagnostic criteria.

Ki67 is a nuclear antigen that is detectable only in proliferating cells and is present at all stages of the cell cycle (G1 – M) but is absent in resting cells in G0. It was identified in 1982 as a proliferation associated biomarker via monoclonal antibody generation in mice immunized with the nuclei of the Hodgkin's Lymphoma cell line L428, with the specific cell cycle characteristics determined two years later. (Schwab et al., 1982, Gerdes et al., 1984). Given the importance of cellular proliferation in tumour development and progression, in addition to its ease of assessment, Ki67 has been studied for prognostic utility in a wide variety of cancers, but has seen limited implementation clinically. In gastroenteropancreatic neuroendocrine neoplasms, Ki67 positivity index forms the basis of a tiered grading system based and is used to predict recurrence and survival in PanNENs (Genç et al., 2018; Klöppel & la Rosa, 2018). In melanoma, Ki67 expression in combination with other IHC based markers has been shown to distinguish melanocytic nevi from tumours, as well as correlate with prognosis in stage I – III disease (Falkenius et al., 2017; Nielsen et al., 2013; Uguen et al., 2018). Most importantly, Ki67 expression (</> 14% positivity) is utilised in breast cancer along with HER2 and ER / PR, to distinguish between two molecular subtypes, Luminal A and Luminal B, which guide subsequent chemotherapeutic regimes (Goldhirsch et al., 2013). Additionally, due to the need to standardise assessment for clinical use, image analysis approaches to Ki67 scoring have been shown to be superior to manual histopathological assessment and highly reproducible between platforms and observers (Acs et al., 2018; Mohammed et al., 2012; Stålhammar et al., 2018).

In the setting of colorectal cancer, Ki67 has been well studied both in the context of primary disease as well as in patients with liver metastasis. In primary CRC, high Ki67 expression has been repeatedly shown to be associated with a better disease specific and overall prognosis (Fluge et al., 2009; Melling, Kowitz, & Simon, 2016; Reimers et al., 2014; C. S. Roxburgh et al., 2013). Interestingly, in CRC liver metastases, this prognostic relationship appears to be inverted, with high Ki67 expression being associated with a poorer

prognosis following colonic resection, and not associated with prognosis following resection of both primary and metastatic disease (Nash et al., 2010b). The association between Ki67 expression and prognosis does appear to depend, however, on the cut off value that is used to stratify patients into high vs low expression groups, with values being determined by various methods ranging from 5% to 60% positivity (Luo et al., 2019).

The aim of the present study is chiefly to investigate automated image analysis approaches to reliably assess Ki67 expression in CRC tissue, and determine its association with survival in independent patient cohorts. Additionally, Roseweir and colleagues proposed a cut off value of 30% Ki67 positivity to stratify patients into high vs low proliferation, therefore, this study also aims to validate this cut off value for prognostic stratification following automated Ki67 expression assessment (Roseweir et al., 2020).

5.2 Materials and Methods

5.2.1 Histopathological Assessment

Immunohistochemical staining of all CRC tissue for Ki67 utilised in this study was conducted according to the methodology described in 2.2.2. Manual histopathological assessment of Ki67 positivity for all cohorts was conducted as described in 2.3.3. Briefly, for TMAs, 100 cells were counted at random and the % positivity was determined for each individual core, where multiple cores were available for a single patient, the % positivity was counted across all cores where sufficient tissue was available and averaged to determine the % positivity for each patient. All histopathological Ki67 assessment was conducted by an observer not involved with the image analysis and scores made available in the respective patient cohort database.

5.2.2 Statistical Analysis

Patient cohort clinicopathological characteristics and test of association with cancer specific / DFS were determined in IBM SPSS version 25 (IBM, New York, USA). P values of <0.05 were considered to demonstrate a statistically significant association between clinicopathological features and clinical outcomes.

All other statistical analyses for this study were conducted in RStudio (RStudio, Boston, MA). Confirmation of cut off values for Ki67 %PI scores were determined by using CSS / DFS as the endpoint with the maxstat and survminer packages. Receiver operator characteristic curves and associated AUC were plotted and calculated using the pROC package. Survival analysis was conducted and associated Kaplan – Meier survival curves produced using the survival, survminer, and survMisc packages, with the log rank statistic used to assess association between Ki67 %PI stratification and CSS / DFS. Hazard ratios and associated 95% confidence intervals for survival analysis were calculated using the Cox proportional hazard model.

5.3 Results

5.3.1 Optimizing Automated Assessment of Ki67 Proliferation Index by Multiple Image Processing Techniques

To determine the Ki67 % Positivity Index (%PI) of CRC tumours, the sections must first be segmented for appropriate tissue morphologies. Given that this study was conducted on TMA cores, a training data set (Training Data Set 3, 2.4.2.3) consisting of manual annotations for four classes: Tumour, Stroma, Necrosis, and Background, was constructed from a single TMA slide from a CRC patient cohort not utilised in the final survival analysis. To conduct tissue segmentation, a U-NET CNN was trained on the data set according to the parameters detailed in 2.4.5.1 and 2.4.5.2, utilising the network performance information and parameters obtained from the H&E TSP algorithm development conducted in Chapter 3, with classifier performance being assessed visually across a range of cores in the initial validation cohort. Representative images of the tissue segmentation algorithm performance are shown in Figure 5.1.

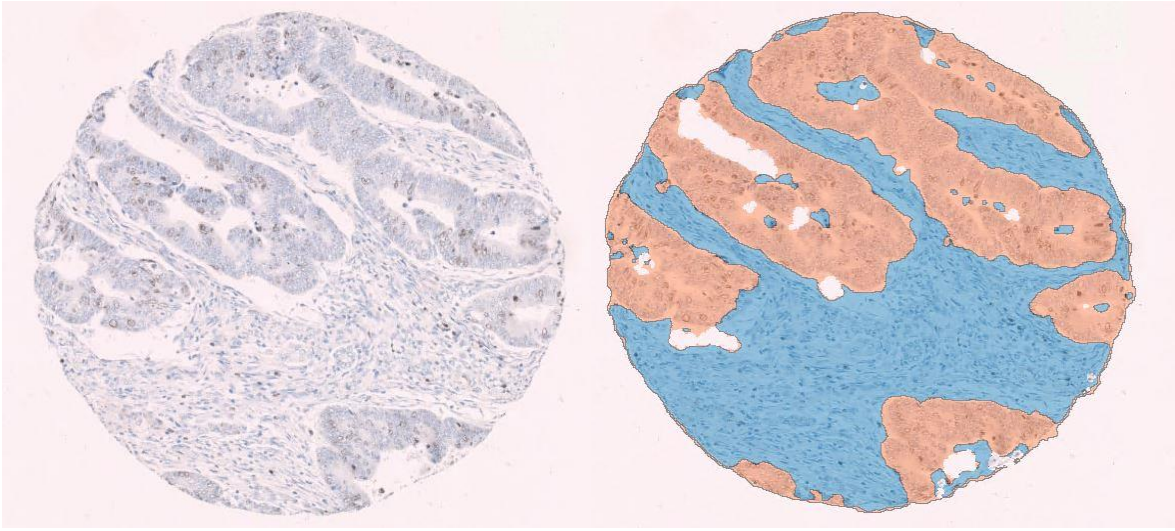


Figure 5.1 Tissue Segmentation of Ki67 IHC TMA Core. *Tissue segmentation algorithm performance for segmenting Ki67 IHC TMA cores into Tumour (orange), Stroma (blue), Necrosis (none present), and Background (white).*

Prior to the tissue segmentation algorithm and subsequent cellular analysis algorithms being applied to the TMA cores, the raw TMA images were de-arrayed using the "Tissuearray" module within the Visiopharm software. An N x N grid was generated for each individual TMA using the TMA maps available for manual histopathological scoring purposes, the grid was then coarsely overlayed on the TMA and each core was finely aligned manually. The de-arrayed cores were then exported back to the "Image Analysis" module

where missing cores and cores with insufficient quality or quantity of tissue for analysis were manually excluded by deleting the ROI but maintaining the image, so the final exported data retained the same coordinate layout as the original TMA. The tissue segmentation algorithm was then run on the TMA core image directed at the ROI surrounding only the remaining viable cores, following which, the tissue class overlays were converted into ROI for subsequent cellular analysis.

In order to determine the optimal method by which to segment and quantify tumour cells stained for Ki67, two cell segmentation algorithms were developed and compared for performance, one utilising traditional image processing methods and one utilising AI. To segment the image via threshold-based processing, the image was initially colour deconvolved into its constituent Haematoxylin and DAB layers, the values of which were multiplied together to generate a single, combined intensity greyscale layer, which was scaled to a range of 0 - 1. A 0 – 1 scaled, 61 x 61 kernel, 7^{th} order Laplacian filter was then applied to the combined intensity layer to distinguish cell boundaries and a threshold value of >0.35 from this filtered layer was coupled with a threshold value of <0.7 applied to the initial combined intensity layer to mark the cell boundaries against the background. Due to the stark increase in intensity of DAB positive cells compared to DAB negative cells, not all negative cells were classified in this initial step, therefore, a 0 – 1 scaled, 71 x 71 kernel, 5^{th} order Laplacian filter was applied to the Haematoxylin colour deconvolved layer and a threshold value of >0.6 was applied to this filtered layer to incorporate the remaining DAB negative cells into the classification. The classified image was then processed to remove any small, misclassified areas (<15µm^2). The cell centroids were isolated from the combined intensity Laplacian filtered layer and expanded to the boundaries detected by the initial classification step to recapitulate the cellular morphology. Following which, a 500µm watershed object separation function with a 6µm wide boundary determined using the combined intensity Laplacian filtered layer heatmap was utilised to distinguish cells from one another at the inferred point of the cellular membrane. Finally, objects <20µm^2 were removed and the remaining cells distinguished as Ki67 positive / negative cells using a threshold of </> 150 applied to the DAB colour deconvolved layer encompassing >5% of the resultant cell objects.

The AI based cell segmentation algorithm was adapted from the "Nuclei Detection, AI (Brightfield)" Analysis Protocol Package (APP) included in the Visiopharm AI Author package. This algorithm was trained on 3 classes: Nuclei, Boundary, and Background from brightfield images of both H&E and IHC sections, with a U-NET CNN utilising unknown training parameters, and no additional training was performed to adapt this algorithm to the

current data set. To categorize detected cells as positive or negative for DAB Ki67 staining, the same threshold function of </> 150 applied to the DAB colour deconvolved layer encompassing >5% of cellular objects utilised in the image processing algorithm was utilised in this algorithm. The output variables from each algorithm were the number of positive and negative cells within the tumour ROI and the Ki67 %PI as calculated by $\left(\frac{N\ Positive\ Cells}{N\ Positive\ Cells + N\ Negative\ Cells}\right) \cdot 100$. The cell segmentation performance of each algorithm was visually assessed across multiple weakly and strongly DAB positive cores to determine both quality of segmentation and determine any colour-based bias towards DAB positive cells. Representative images of the performance of both algorithms are shown in Figure 5.2.

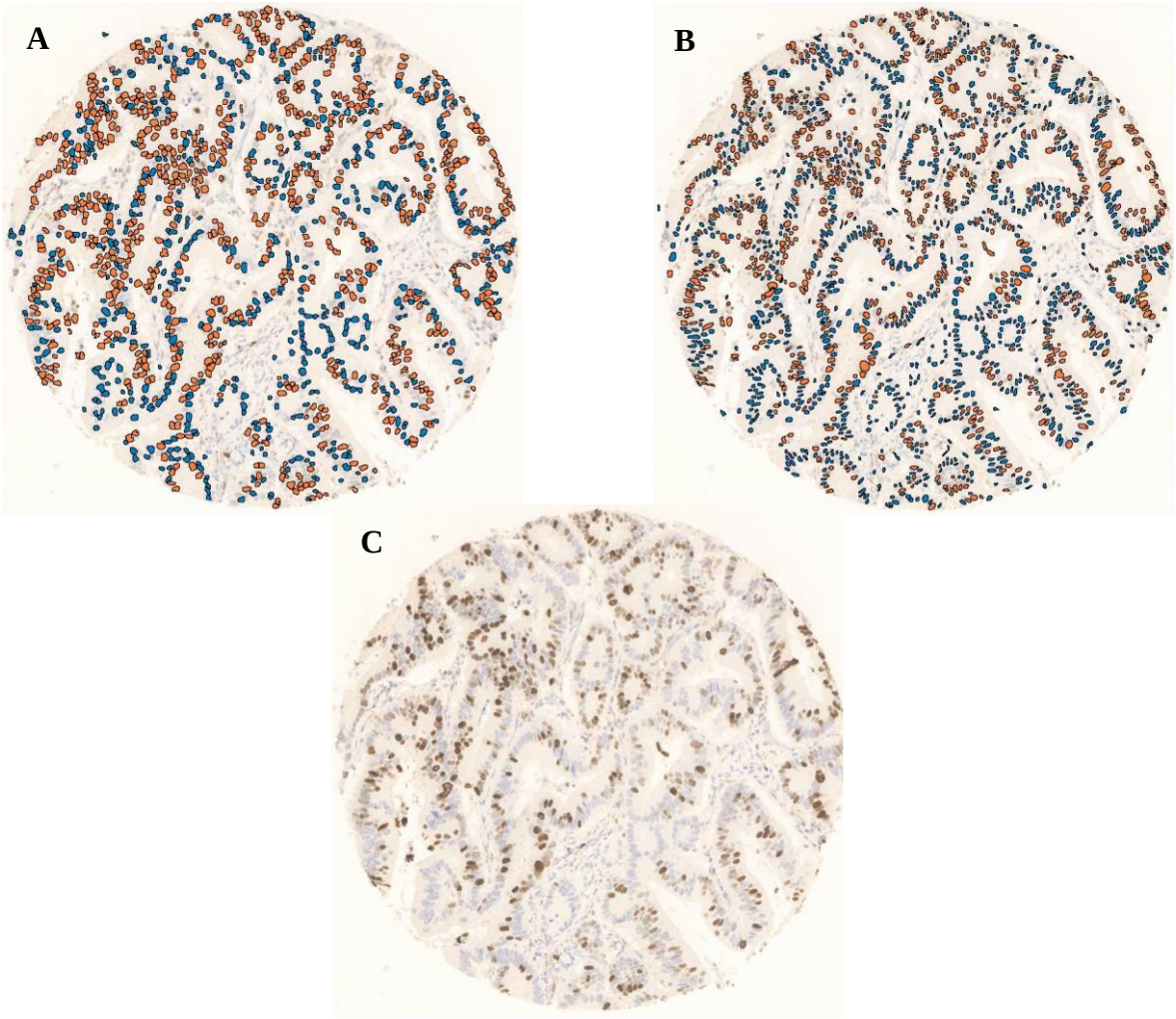


Figure 5.2 Cell Segmentation Algorithms on Ki67 IHC. *Performance of image processing (A) and AI (B) based cell segmentation algorithms on Ki67 stained CRC tissue (C).*

To assess the performance of each approach against manual histopathological assessment, a TMA constructed from the Glasgow Development Cohort (clinicopathological features detailed in 3.3.2) stained for Ki67, was initially classified using the AI-based tissue segmentation algorithm described above and the class overlays converted to ROI. The tumour ROI generated by this algorithm were then analysed for Ki67 %PI using both cell segmentation algorithms on all viable cores in the TMA and the average of the cores taken to generate a Ki67 %PI for each patient, the final workflow for this algorithm is shown in Figure 5.3. The scores were compared to manual histopathological assessment via ROC and Intraclass Correlation Coefficient (ICC) analysis. The threshold – based algorithm generated an AUC = 0.926 and an ICC3 = 0.78 (95% CI 0.73 – 0.83), and the AI – based algorithm produced an AUC = 0.947 and an ICC3 = 0.83 (95% CI 0.78 – 0.86) (Figure 5.4).

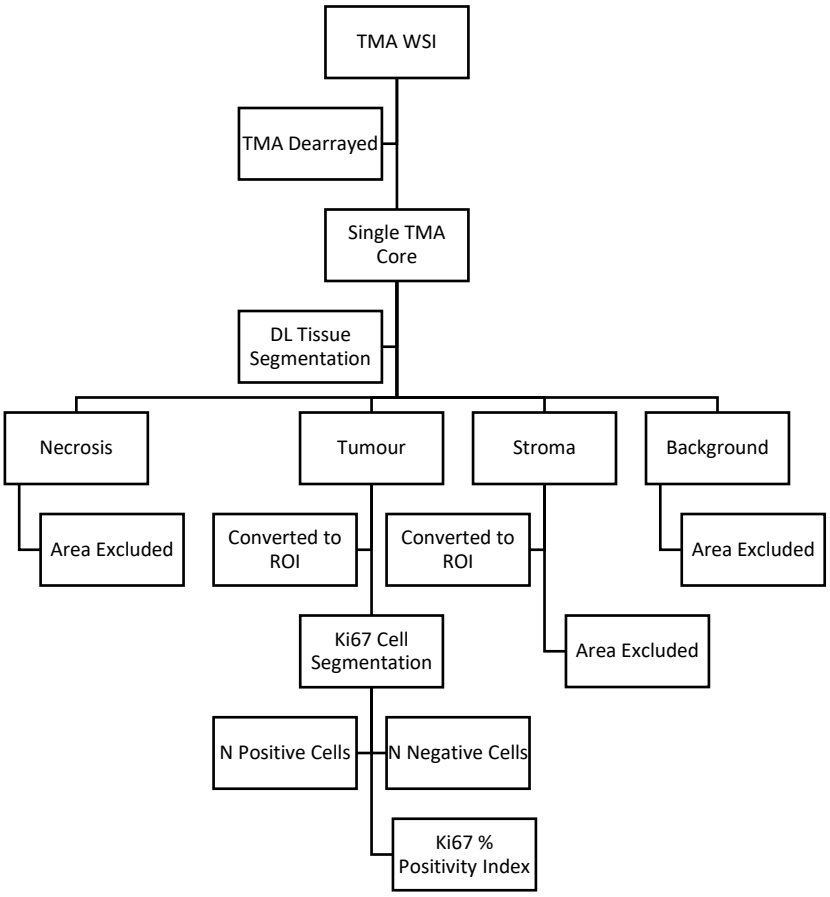


Figure 5.3 Ki67 % Positivity Index Algorithm Workflow. *Workflow to determine Ki67 %PI from IHC TMA WSI. Workflow is identical between image processing and AI – based algorithms.*

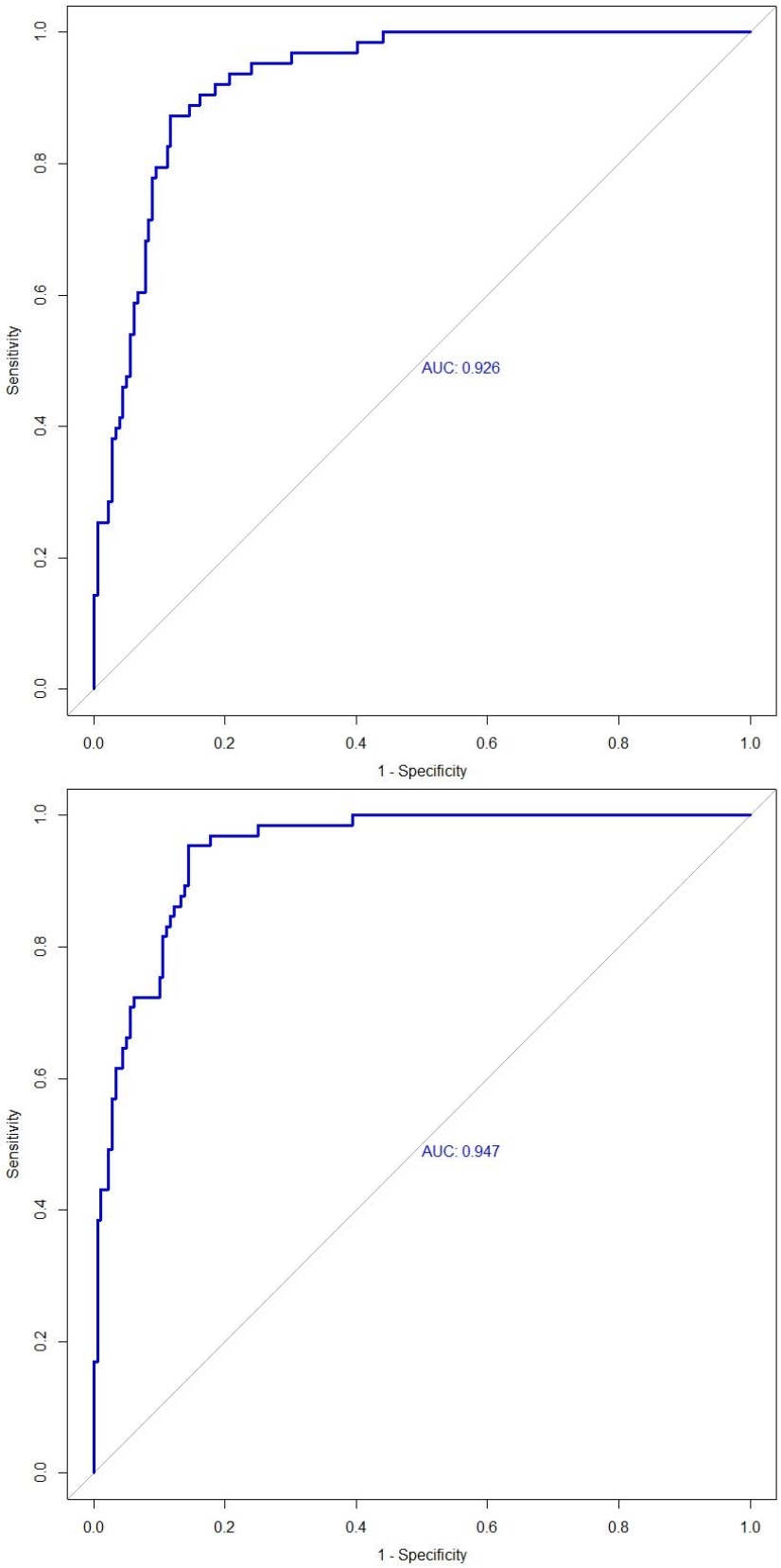


Figure 5.4 Receiever Operator Characteristic of Ki67 %PI Algorithms in Glasgow Development Cohort. *ROC analysis of Ki67 %PI scores generated by an image processing (A) and an AI – based (B) cell segmentation algorithm assessed against histopathological Ki67 %PI scores with 30% PI cut off in Glasgow Development Cohort. Image processing algorithm AUC = 0.926 and ICC3 = 0.78. AI – based algorithm AUC = 0.947 and ICC3 = 0.83.*

Bland – Altman plots were constructed for each image analysis algorithm against the histopathological Ki67 %PI scores to assess the relationship between the image analysis and manual approaches, and to determine whether any systemic bias is present in either image analysis algorithm (Figure 5.5).

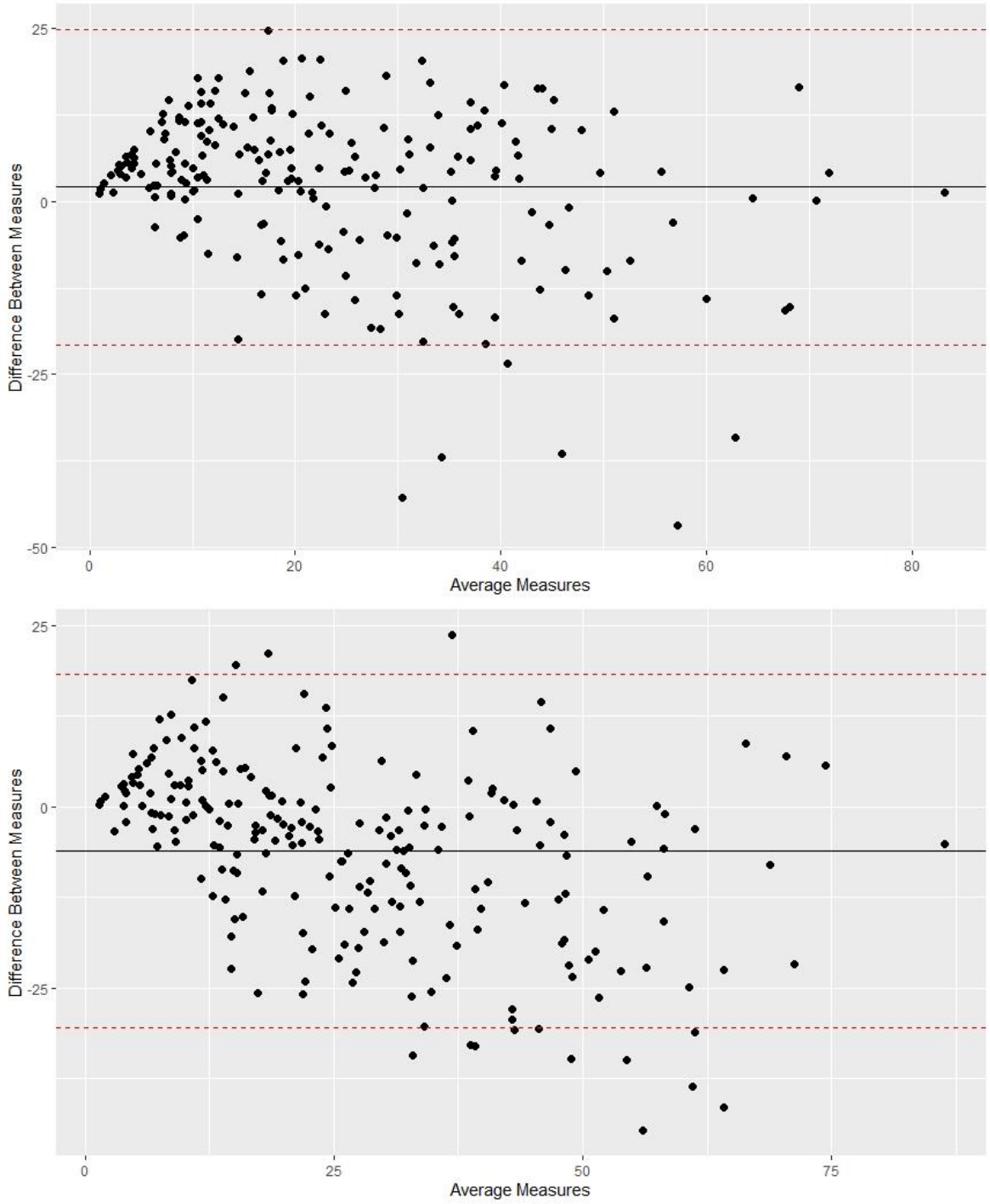


Figure 5.5 Relationship between histopathological and image analysis determined Ki67 %PI in Glasgow Development Cohort. *Bland – Altman plots demonstrating the relationship and bias between Ki67 %PI determined by manual histopathological assessment and an image processing (A) and an AI – based (B) cell segmentation algorithm in the Glasgow Development Cohort.*

The Ki67 %PI scores generated by each algorithm were assessed for prognostic performance against CSS in the Glasgow Development cohort. Initially, the cut off values for the scores generated by each algorithm were determined using the maxstat and survminer packages in R against the CSS for each patient. Given that the cut off value utilised in all histopathological Ki67 survival analysis within our lab is 30%, the aim of generating a cut off for each algorithm was to determine their respective distances from this value and assess the utility of the 30% cut off for CSS prognosis by each algorithm. The cut off values determined for the image processing and AI – based algorithms were 18.65% and 29.57% respectively (Figure 5.6).

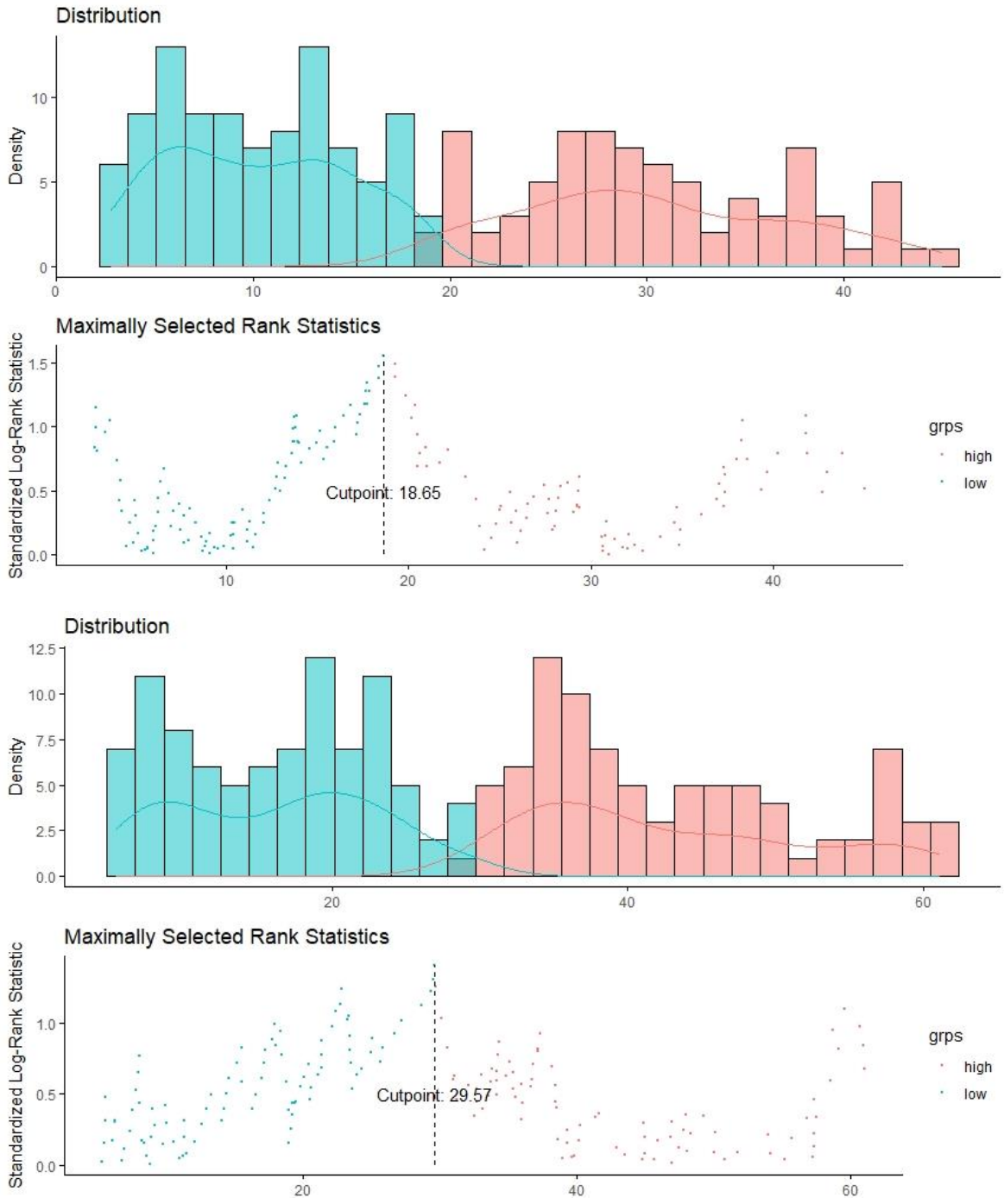


Figure 5.6 Determining Optimal Cut Off Values For Ki67% PI Cell Segmentation Algorithms In Glasgow Development Cohort. *Distribution and optimal cut off values of Ki67 %PI scores determined by an image processing (A) and AI – based (B) cell segementation algorithm for CSS in the Glasgow Development Cohort. The cut off values for the image processing and AI – based algorithms in this cohort were 18.65% and 29.57% respectively.*

Kaplan – Meier survival analysis was performed to assess the association between CSS and the two image analysis approaches. The Ki67 %PI scores generated by each algorithm were used to stratify patients into high vs low Ki67 expression utilising the 30% PI cut off value determined via manual histopathological assessment. Histopathological Ki67 assessment in this cohort showed a non – statistically significant association with CSS but an increased hazard ratio for low Ki67 expression (HR 1.34 95% CI 0.7827 – 2.294, *P = 0.29*, Figure 5.7). Ki67 expression determined by the image processing algorithm showed a slightly increased hazard ratio for CSS but did not show a statistically significant association (HR 1.136 95% CI 0.6933 – 1.861 *P = 0.61*, Figure 5.7), whereas the AI – based algorithm showed a comparable association with CSS to the manual histopathological assessment (HR 1.236 95% CI 0.7681 – 2.1.989, *P = 0.38*, Figure 5.8).

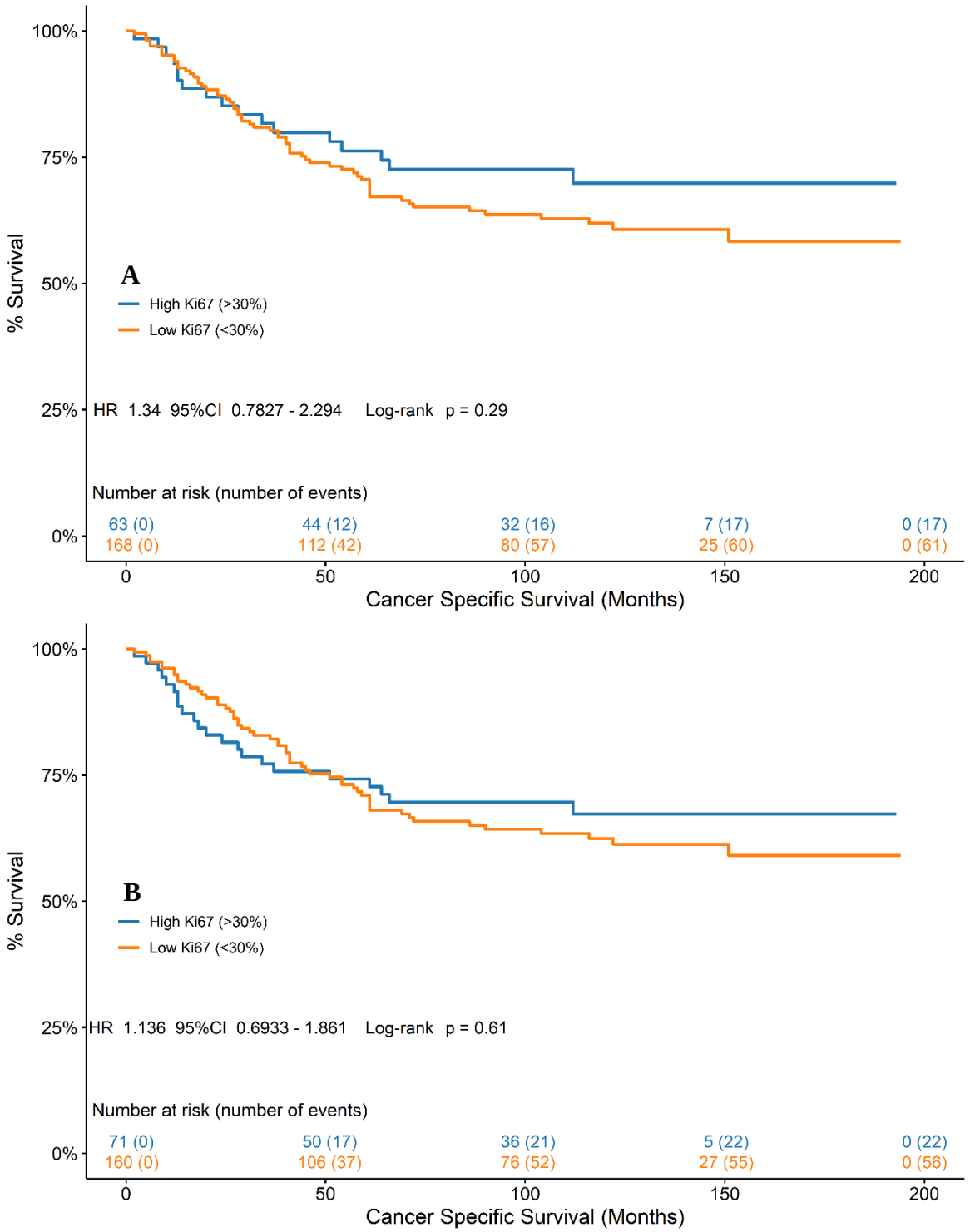


Figure 5.7 Relationship Between Image Processing Determined Ki67 Expression and CSS in Glasgow Development Cohort. *Association between CSS and Ki67 %PI determined by manual histopathological assessment (A) and an image processing cell segmentation algorithm (B) in the Glasgow Development Cohort.*

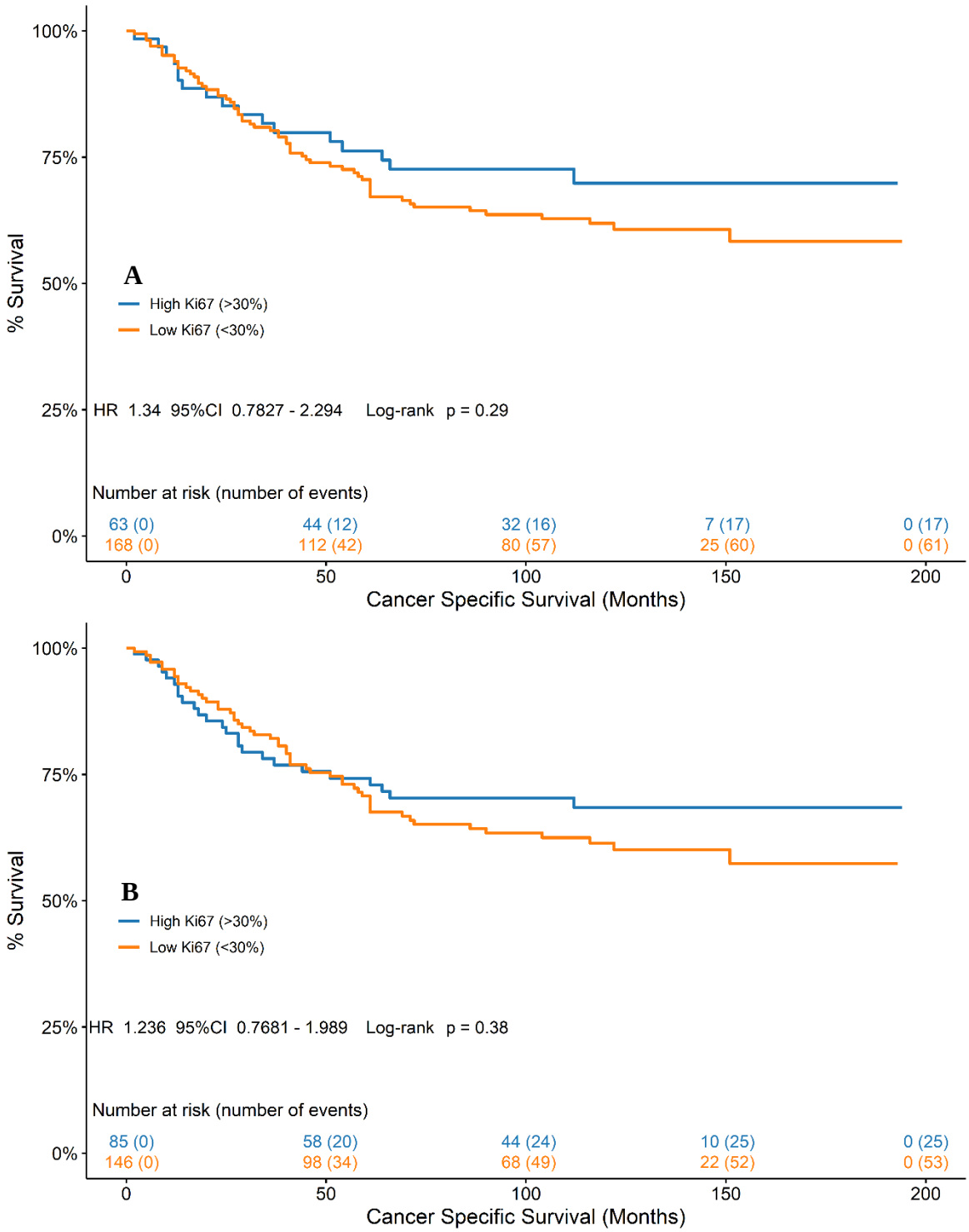


Figure 5.8 Relationship Between AI Determined Ki67 Expression and CSS in Glasgow Development Cohort. *Association between CSS and Ki67 %PI determined by manual histopathological assessment (A) and an AI – based cell segmentation algorithm (B) in the Glasgow Development Cohort.*

5.3.2 Validation of Ki67 Cell Segmentation Algorithms in The Glasgow Validation Cohort

As detailed in 5.3.1, the AI – based Ki67 cell segmentation algorithm demonstrated improved performance over the traditional image processing algorithm in terms of faithful cell segmentation (Figure 5.2), statistical measures of agreement with manual histopathological assessment (Figure 5.4), and in terms of association with CSS (Figure 5.8). To confirm these findings, both algorithms underwent the same analysis in the Glasgow Validation Cohort (clinicopathological features detailed in 3.3.2, Table 3.2), which had previously been histopathologically assessed for Ki67 expression by an independent observer (5.2.1). The application of the algorithms to this cohort followed an identical workflow to that used in the initial validation in the Glasgow Development Cohort (Figure 5.3) and was applied over 4 cores per patient, with the final Ki67 %PI for each patient determined by averaging the score of all viable cores in the TMA. The Ki67 %PI for each patient from each algorithm initially underwent ROC analysis and ICC analysis against the histopathological Ki67 assessment stratified and percentage scores respectively to determine the statistical agreement. In this cohort, the threshold based algorithm produced an AUC = 0.928 and ICC3 = 0.74 and the AI – based cell segmentation algorithm produced an AUC = 0.929 and an ICC3 of 0.83 (Figure 5.9).

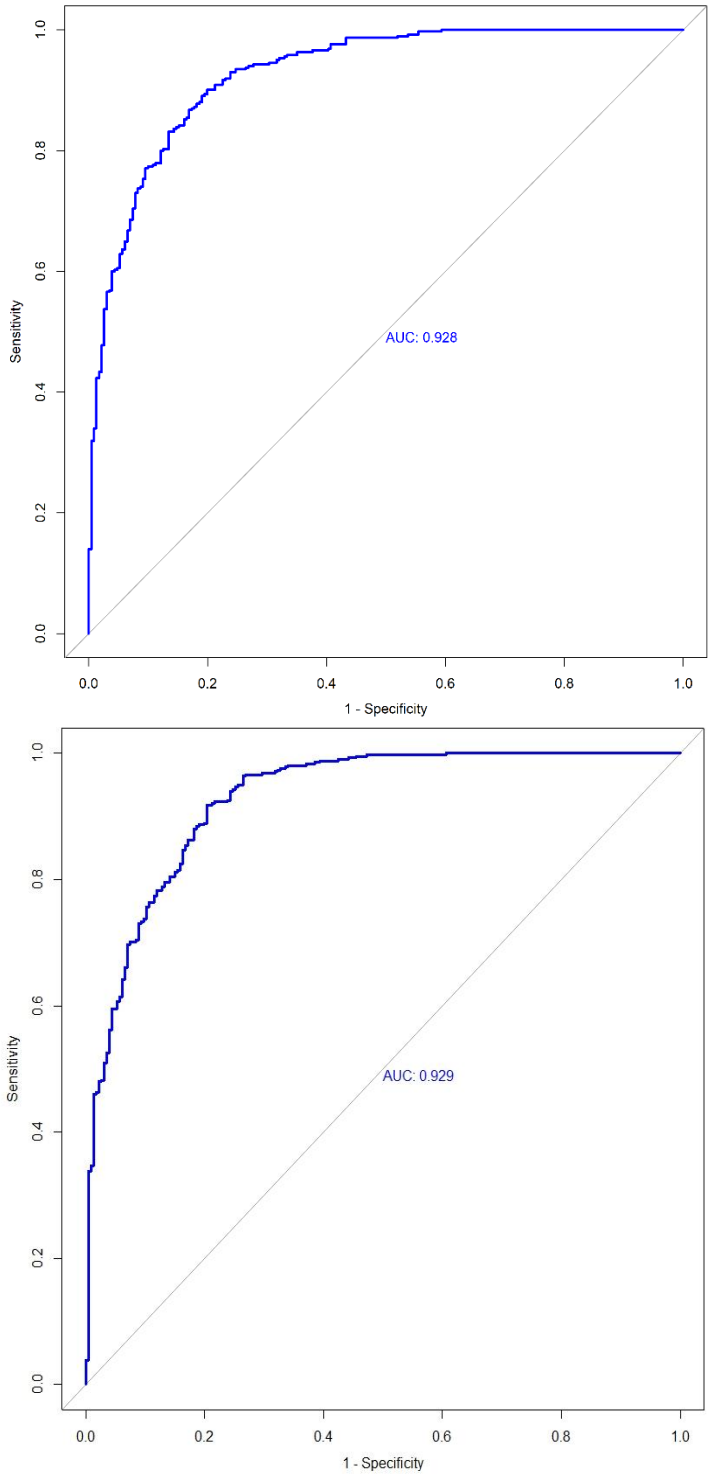


Figure 5.9 Receiver Operator Characteristic of Ki67 %PI in the Glasgow Validation Cohort. *ROC curve of Ki67 %PI determined by a threshold – based (top) AI – based (bottom) cell segmentation algorithm against manual histopathological assessment. Threshold AUC = 0.928 and ICC3 = 0.74, AI – based AUC = 0.929 and ICC3 = 0.83.*

Bland – Altman plots were constructed to assess the relationship and determine any systemic bias between the cell segmentation algorithms' generated Ki67 %PI and the manual histopathological assessment (Figure 5.10).

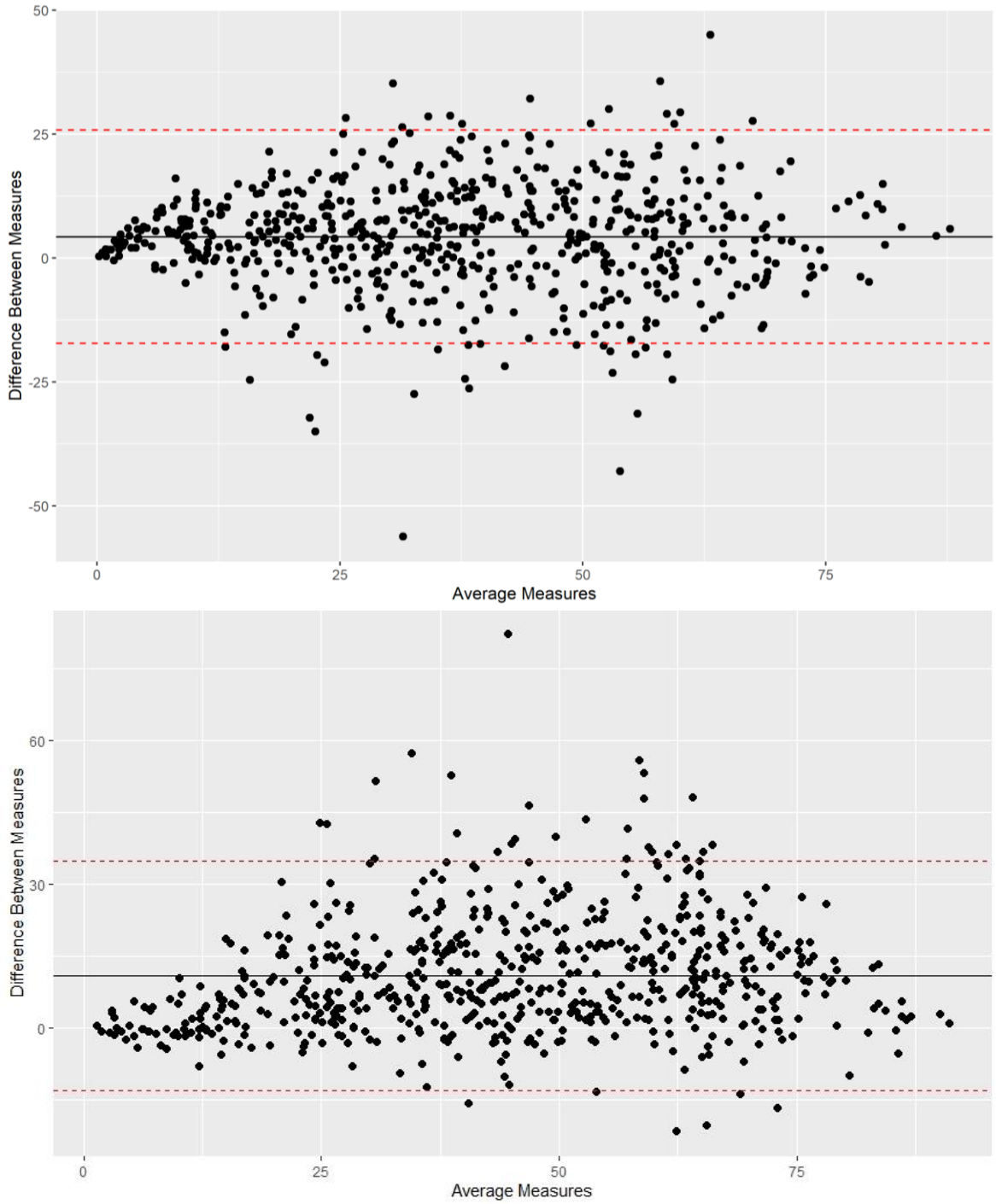


Figure 5.10 Relationship Between Histopathological and Image Analysis Determined Ki67 %PI in the Glasgow Validation Cohort. *Bland – Altman plots demonstrating the relationship and bias between histopathological and threshold – based cell classification algorithm (top) and AI – based cell classification (bottom) determined Ki67 %PI in the Glasgow Validation Cohort.*

Cut off values for the Ki67 %PI scores determined by the algorithms for CSS were generated using the maxstat and survminer packages in R to determine the level of discrepancy with the 30% cut off value used in manual histopathological assessment. The CSS cut off value for the threshold algorithm generated scores in this cohort was 59.37% and 34.51% for the AI – based algorithm (Figure 5.11).

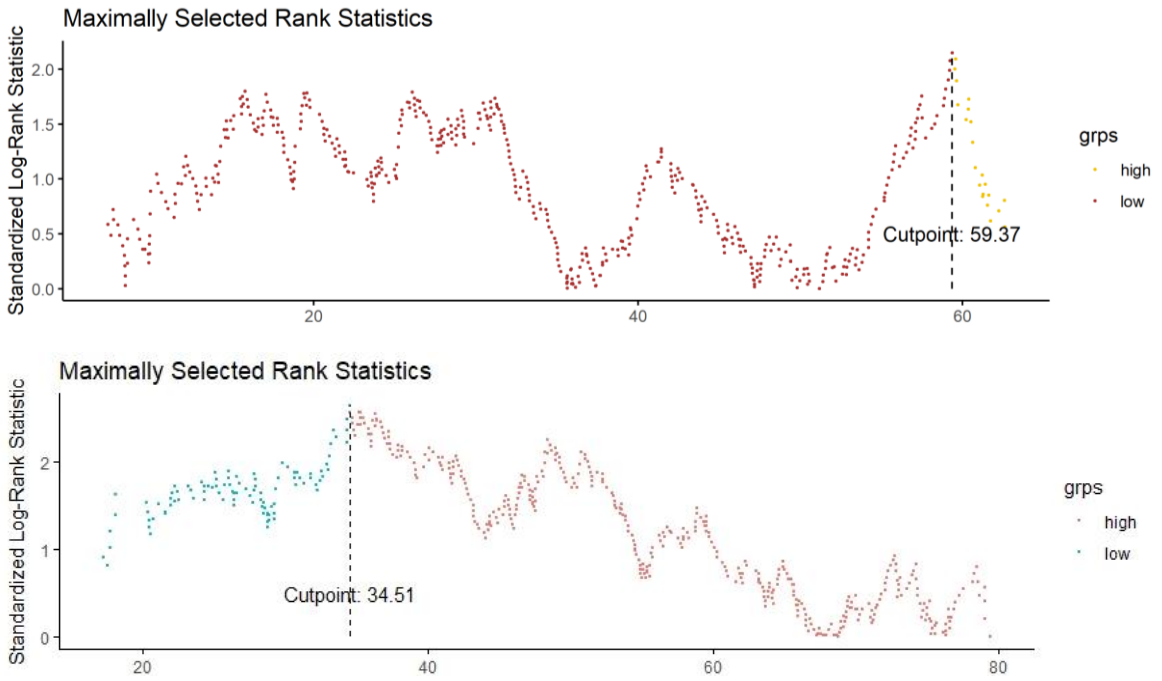


Figure 5.11 Determining an optimal cut off value for CSS by Ki67 %PI determined by image analysis cell classification algorithms in Glasgow Validation Cohort. *Distribution of AI – based image analysis determined Ki67 %PI scores and optimal cut off value for high vs low expression based on CSS in the Glasgow Validation Cohort. For this algorithm in this cohort, the optimal cut off value was 59.37% for the threshold algorithm and 34.51% for the AI – based algorithm.*

The algorithm generated scores were used to stratify the patients into high vs low Ki67 expression using the </> 30% cut off value utilised in manual histopathological assessment. Kaplan – Meier analysis was performed on the stratified scores to assess the relationship with CSS and compare survival stratification to manual histopathological assessment. The AI – based algorithm generated Ki67 %PI showed a stronger statistical association with CSS in this cohort than the threshold algorithm determined indices, which failed to reach significance, HR 1.473 95% CI 1.066-2.036 *P = 0.018* and HR 1.271 95% CI 0.9321 – 1.734 *P = 0.18*, respectively (Figure 5.12 & Figure 5.13). Additionally, the AI – based algorithm outperformed manual histopathological assessment for CSS stratification (Figure 5.13).

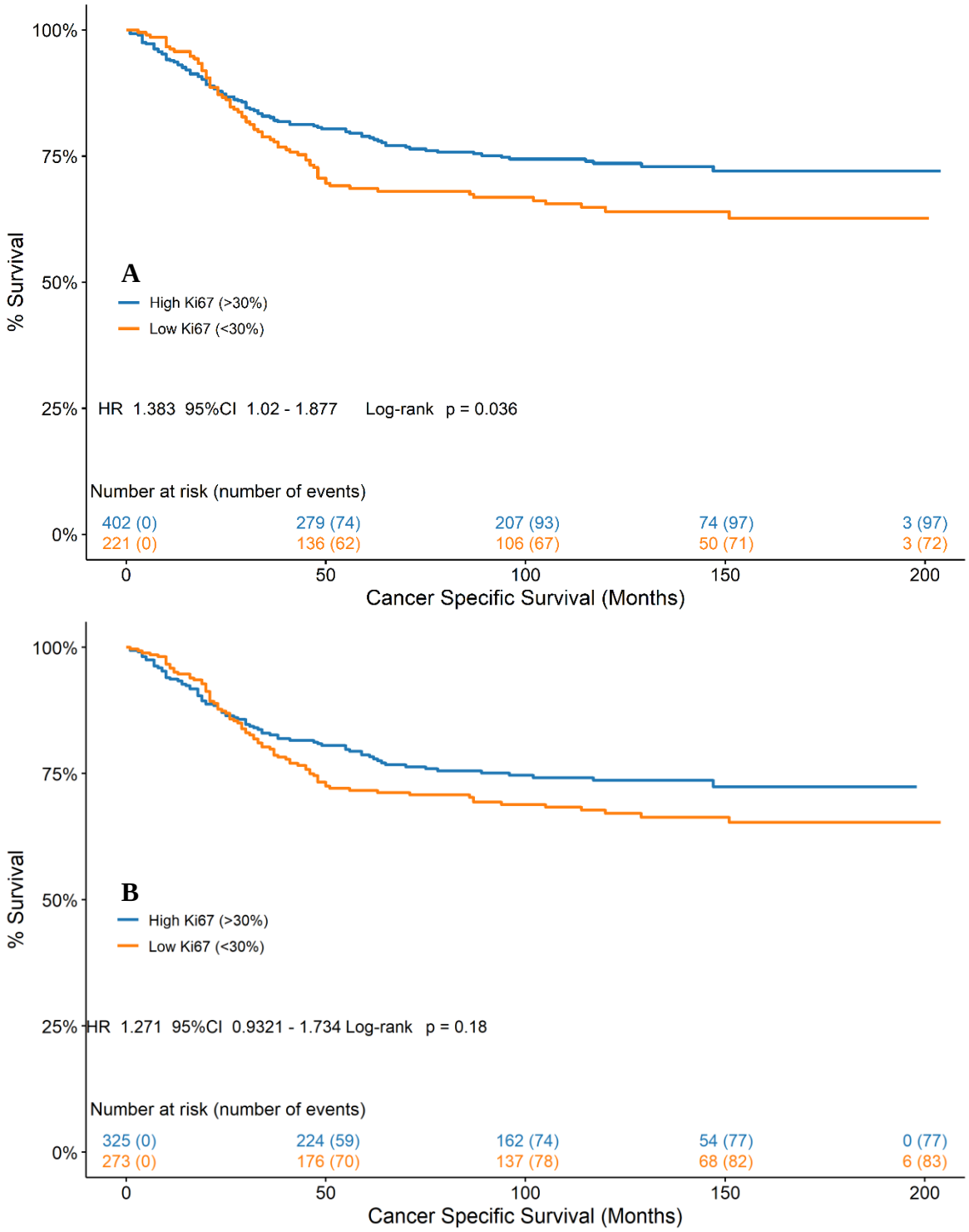


Figure 5.12 Relationship between Ki67 %PI and CSS in Glasgow Validation Cohort. *Relationship between CSS and Ki67 %PI determined by manual histopathological assessment (A) and threshold – based image analysis (B) in the Glasgow Validation Cohort.*

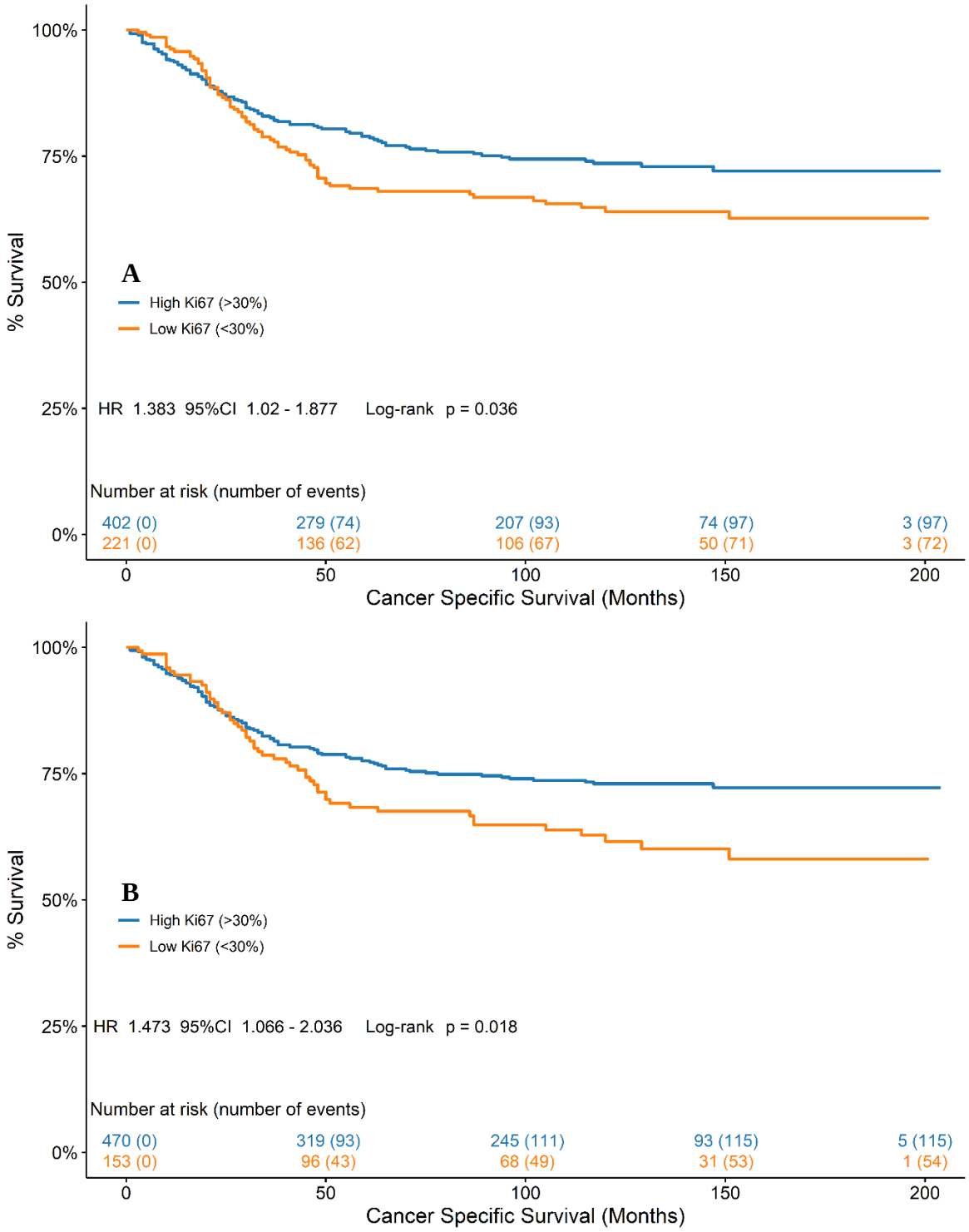


Figure 5.13 Relationship between Ki67 %PI and CSS in Glasgow Validation Cohort. *Relationship between CSS and Ki67 %PI determined by manual histopathological assessment (A) and AI – based image analysis (B) in the Glasgow Validation Cohort.*

5.3.3 Automated Ki67 Expression Assessment by Image Analysis Predicts Disease Specific Survival in a Previously Unscored Patient Cohort

Validation of the two image analysis approaches in the Glasgow Validation cohort confirmed the superior performance of the AI – based algorithm over the threshold algorithm demonstrated in the Glasgow Development cohort. The AI – based algorithm showed a greater association with CSS than both the threshold algorithm and manual histopathological assessment (Figure 5.12 & 5.13), and again produced a cut off value close to the 30% determined manually (Figure 5.11), whereas the threshold produced inconsistent thresholds significantly above and below the accepted value (Figure 5.6 & 5.11). The AI – based algorithm was therefore selected as the most appropriate algorithm to conduct novel analyses in subsequent cohorts. As stated in previous chapters, it is vital to the clinical translatability of image analysis algorithms that their application to image data from novel patient cohorts retains performance in terms of both segmentation and clinical outcomes. Development and validation of the automated Ki67 approaches was conducted on well – established patient cohorts that had previously been histopathologically scored for Ki67 %PI. Therefore, to validate image analysis Ki67 assessment in a completely novel setting, the same workflow used in the Glasgow Development and Validation patient cohorts was applied to tissue from a cohort recently established at the Glasgow Royal Infirmary hospital, that had not been previously scored for Ki67.

The Glasgow Royal Infirmary (GRI) cohort comprised 787 stage I – III colorectal cancer patients undergoing potentially curative resection at the GRI hospital between 1997 – 2012. Patients were predominantly male (55.2%) over the age of 65 (62.3%) and presenting with primarily T – Stage III (56.9%), node negative disease (60.3%) with a relatively even distribution of right, left, and rectal disease (38.5%, 31.1%, 30.4%, respectively). Patients were excluded from final survival analysis if they were administered neoadjuvant therapy, died within 30 days of surgery, or presented with Stage IV disease, and if there was insufficient tissue for image analysis. Following exclusion criteria, 638 patients remained and were included in the final analysis. The primary clinical outcome was CSS defined as the time from the date of surgery to the date of death from primary disease. Clinicopathological characteristics associated with CSS were Age ($P = 0.012$), T Stage ($P < 0.0001$), and N Stage ($P < 0.0001$) (Table 5.1).

Clinicopathological Characteristic	N (%)	Clinical Outcome Significance
Age		**0.012**
<65	296 (37.7)	
>65	489 (62.3)	
Sex		0.201
Female	352 (44.8)	
Male	433 (55.2)	
T Stage		**<0.0001**
I	34 (4.3)	
II	84 (10.7)	
III	447 (56.9)	
IV	219 (27.9)	
N Stage		**<0.0001**
0	473 (60.3)	
1	223 (18.5)	
2	89 (7.4)	
Tumour Site		0.259
Right	302 (38.5)	
Left	244 (31.1)	
Rectal	239 (30.4)	
Differentiation		*0.234*
Well	23 (1.9)	
Moderate	681 (86.5)	
Poor	72 (6.0)	
Adjuvant Chemotherapy		0.744
Yes	200 (25.5)	
No	583 (74.5)	

Table 5.3 Clinicopathological Characteristics of Glasgow Royal Infirmary Cohort. *Number (%) of patients with clinicopathological characteristic and association with CSS in GRI patient cohort.*

An identical workflow used in the previous cohort TMAs was applied to the GRI cohort TMA, composed of three cores per patient from which the resultant Ki67 %PI scores were averaged. The scores were then analysed using the survminer and maxstat packages in R to generate a cut off value based on CSS and determine the level of discrepancy with the

30% cut off used in manual histopathological assessment. The cut off value determined for this cohort with the image analysis generated Ki67 scores was 30.4% (Figure 5.14).

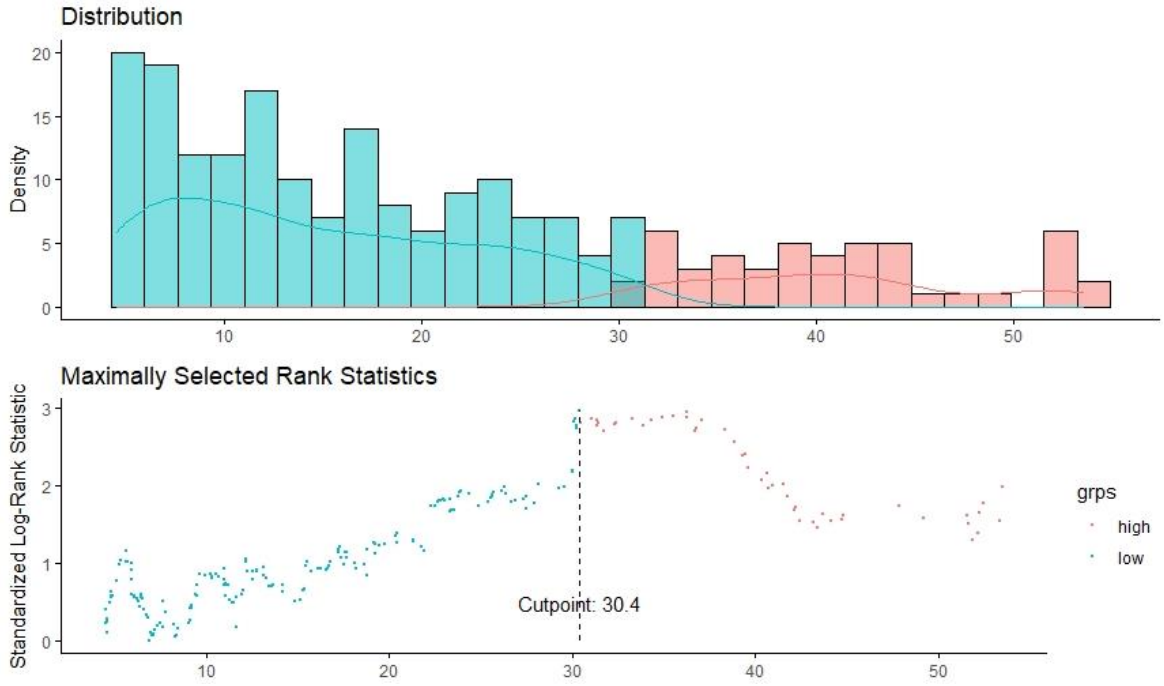


Figure 5.14 Determining an optimal cut off value for CSS by Ki67 %PI determined by AI – based image analysis algorithm in GRI Cohort. *Distribution of AI – determined Ki67 %PI and optimal cut off value for high vs low expression based on CSS in the GRI patient cohort. The cut off value for this cohort is 30.4%.*

The algorithm generated scores were used to stratify patients into high vs low expression based on the 30% cut off value used in manual histopathological assessment and validated by the maximally selected rank statistics analysis in Figure 5.14, following which, the stratified scores underwent Kaplan – Meier survival analysis to determine the association between the Ki67 %PI and CSS. The stratified AI – determined Ki67 scores were significantly associated with CSS in this previously unscored cohort (HR 1.626 95% CI 1.125 – 2.352 $P = 0.009$, Figure 5.15).

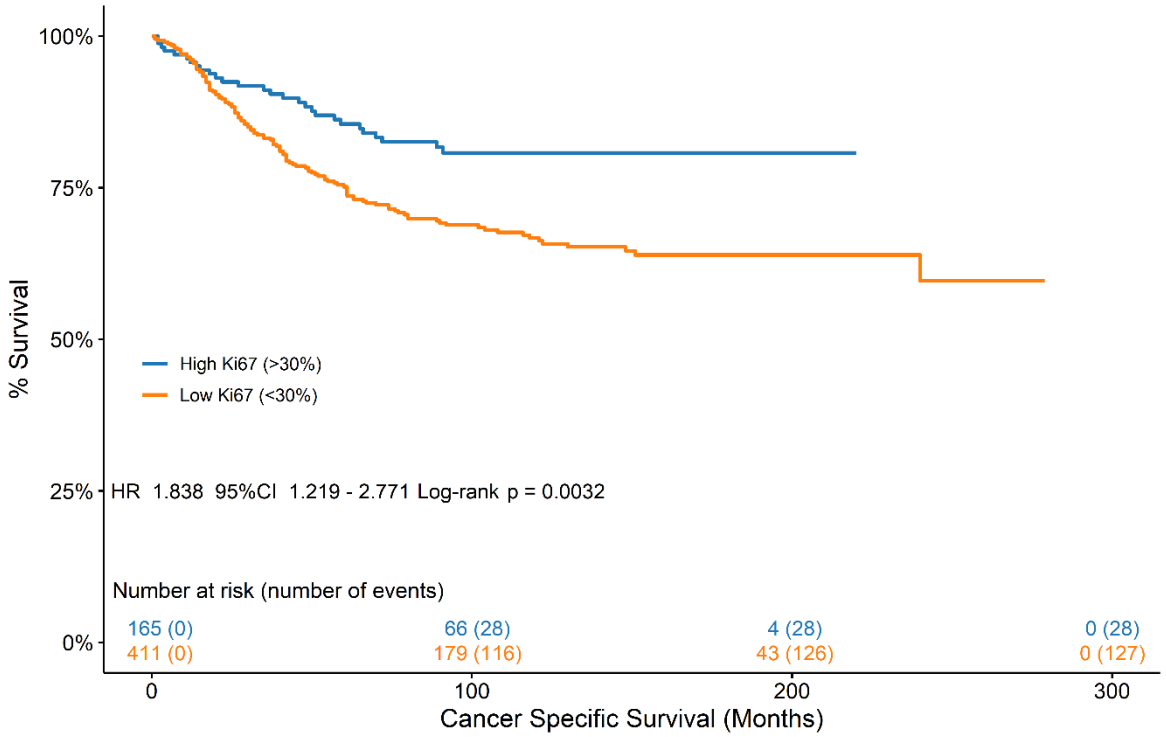


Figure 5.15 Relationship Between CSS and AI – Determined Ki67 %PI in GRI Cohort. *Association between CSS and AI – determined Ki67 expression in the Glasgow Royal Infirmary patient Cohort.*

Univariate and multivariate cox regression was performed to assess the relationship between the Ki67 %PI determined via image analysis, clinicopathological features, and CSS (Table 5.2). On multivariate analysis, Ki67 index was found to be significantly associated with CSS (HR 0601 95%CI 0.411-0.879, $p = 0.009$), independent of Age ($p = 0.003$), N-Stage ($p < 0.001$), and Tumour Budding ($p = 0.001$).

Univariate				Multivariate			
	HR	95% CI	p-value		HR	95% CI	p-value
Age				**Age**			
<65	1.0			<65	1.0		
65-75	0.942	0.636-1.396	0.766	65-75	0.975	0.634-1.499	0.908
>75	1.605	1.119-2.302	**0.01**	>75	1.826	1.228-2.714	**0.003**
Sex							
Female	1.0						
Male	1.320	0.969-1.797	0.078				
T Stage				**T Stage**			
1	1.0			1	1.0		
2	1.362	0.375-4.950	0.639	2	1.173	0.251-5.473	0.840
3	2.690	0.851-8.500	0.092	3	2.034	0.494-8.380	0.326
4	5.166	1.624-16.431	**0.005**	4	3.002	0.718-12.548	0.132
N Stage				**N Stage**			
0	1.0			0	1.0		
1	2.521	1.801-3.528	**<0.001**	1	1.877	1.296-2.718	**<0.001**
2	4.137	2.736-6.254	**<0.001**	2	2.825	1.781-4.482	**<0.001**
Site							
Right	1.0						
Left	1.021	0.714-1.459	0.909				
Rectal	1.138	0.784-1.650	0.497				
Differentiation				**Differentiation**			
Well	1.0			Well	1.0		
Moderate	3.143	0.779-12.686	0.108	Moderate	3.450	0.850-14.006	0.083
Poor	4.728	1.101-20.300	**0.037**	Poor	3.364	0.754-14.996	0.112
Tumour Budding				**Tumour Budding**			
Low	1.0			Low	1.0		
High	1.871	1.370-2.555	**<0.001**	High	1.752	1.254-2.447	**0.001**
Ki67 PI				**Ki67 PI**			
Low	1.0			Low	1.0		
High	0.539	0.373-0.778	**<0.001**	High	0.601	0.411-0.879	**0.009**

Table 5.2 Relationship between Ki67 PI, Clinicopathological Features and CSS in GRI Cohort. Hazard ratios and 95% confidence intervals determined via Cox proportional hazards regression and multivariate analysis conducted using backwards conditional method.

5.3.4 Automated Assessment of Ki67 Expression at the Invasive Margin and Tumour Core in a Clinical Trial Cohort

Following validation of automated Ki67 expression assessment using the AI – based cell classification algorithm in the previously scored Glasgow Development and Validation cohorts, and the previously unscored GRI cohort, the algorithm was further validated in the TransSCOT clinical trial cohort (detailed in 3.3.5, Table 3.4). The TransSCOT cohort comprises 2913 patients with full section data, from which 1317 sections were immunohistochemically stained for Ki67 to be utilised in a study validating the Phenotypic Subtypes and its components for prognostic significance. The Ki67 full sections were then analysed in a single, pre – determined field of view at 400x magnification using an automated hotspot – based positive cell counter that was part of the SlidePath (Leica, UK) digital image hub (Roseweir et al., 2020). A secondary aim of the full section analysis was to ensure that areas selected to be taken for a TMA constructed from this cohort were biologically representative of the whole tumour. The TMA was constructed from 2352 of the 2913 patients utilising two 0.8mm cores per patient, one from the Tumour Core (TC) and one from the Invasive Margin (IM). Following exclusion of patients with missing data and those that died within 30 days of surgery, 2079 patients with both IM and TC data remained for survival analysis. In order to derive as much clinically relevant data as possible from the cohort, the TMA was selected as the mode of analysis for Ki67 expression and subsequently stained via IHC.

It has been previously shown that there is a marked difference in expression levels and prognostic effect of Ki67 assessed at the luminal edge / tumour core and invasive margin. A study investigating proliferation in Dukes' B tumours demonstrated that proliferative activity was significantly higher at the luminal border compared to the invasive margin, and that low proliferative activity at the invasive margin was correlated with a worse prognosis but no correlation with survival was seen when Ki67 was assessed at the luminal border (Palmqvist et al., 1999). A later study investigating apoptosis and proliferation in adenoma and carcinoma confirmed that proliferative activity was higher in the central part of the tumour compared to the invasive margin in CRC. Additionally, this study showed a trend towards significantly increased proliferative activity at the IM in the right vs left colon (Hörkkö & Mäkinen, 2003). Therefore, to further investigate this phenomenon of difference in expression and its correlation with prognosis in this cohort, the AI – based cell classification algorithm, previously validated in three CRC cohorts, was applied to both IM and TC cores from each patient using an identical workflow to the validation studies. The

Ki67 %PI scores generated for each patient at the IM and TC were compared in terms of expression level and the effect on DFS.

To determine the discrepancy with the 30% cut off value used in histopathological Ki67 assessment, an optimal cut off value for the Ki67 %PI scores generated by the AI cell classification algorithm was generated against DFS status, for both the IM and TC for each patient, using the survminer and maxstat packages in R studio. The optimal cut off values determined for DFS, for the IM and TC, in this cohort were 13.41% and 10.16%, respectively (Figure 5.16).

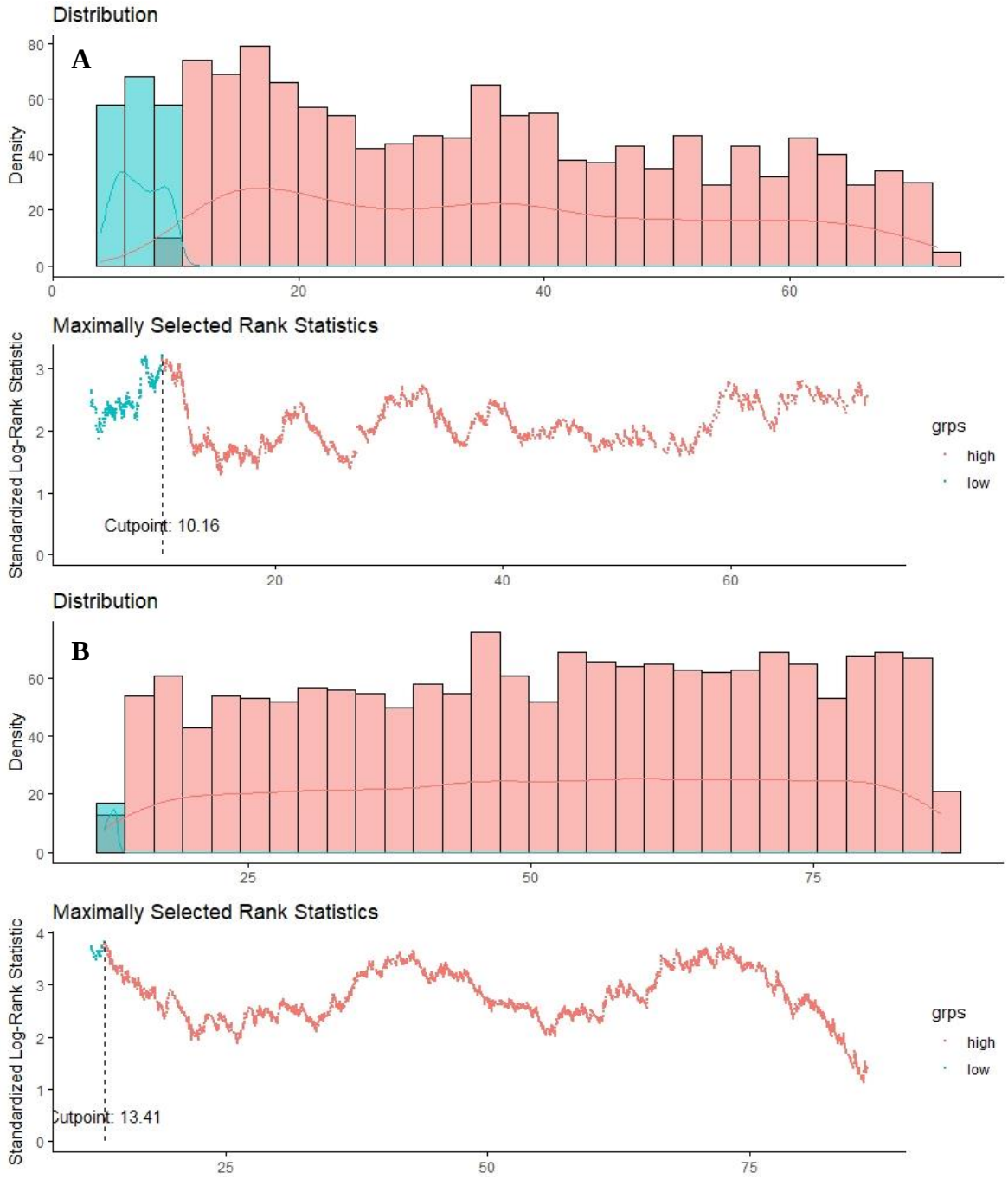


Figure 5.16 Determining Optimal Cut Off Values For Ki67 in TransSCOT Cohort. *Distributions and optimal cut off values for Ki67 %PI determined by AI cell classification at the invasive margin (A) and tumour core (B) in the TransSCOT cohort.*

The AI algorithm generated scores were used to stratify patients into high vs low expression based on the 30% cut off value used in manual histopathological assessment, and Kaplan – Meier analysis was performed to assess the association of the stratified scores with DFS at the invasive margin and tumour core. In both cases, high Ki67 expression (>30% PI) conferred a better DFS prognosis however, Ki67 expression at the invasive margin showed a greater hazard ratio and correlation with DFS in this cohort than within the tumour core, correlating with previous studies that Ki67 expression carries greater prognostic significance at the invasive margin (IM = HR 1.281 95% CI 1.085-1.513 $P = 0.0035$, TC = HR 1.236 95% CI 1.032-1.481 $P = 0.021$, Figure 5.17).

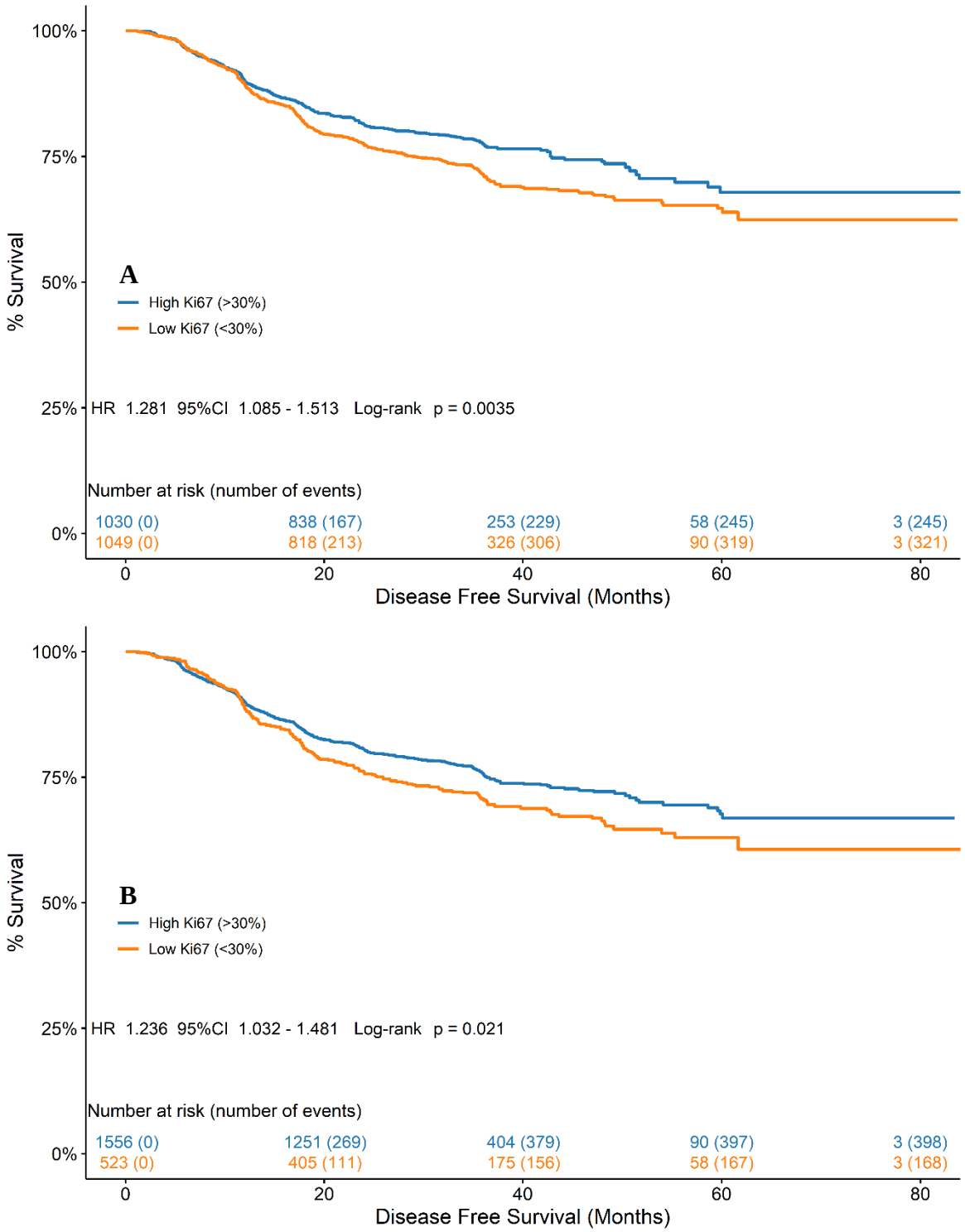


Figure 5.17 Relationship Between Ki67 Expression and DFS in TransSCOT Cohort. *Association between AI – determined Ki67 expression and DFS at the invasive margin (A) and tumour core (B) in the TransSCOT clinical trial cohort.*

To evaluate any potential difference in Ki67 expression levels between the invasive margin and tumour core, and if sidedness influenced expression, a Wilcoxon matched pairs signed – rank test was performed on the paired Ki67 %PI scores from the IM and TC cores for each patient grouped by the disease site. In both left and right sided disease, there was significantly higher Ki67 expression in the tumour core, further correlating with previous studies, with left – sided disease showing a slightly greater difference in expression level than right – sided disease ($P < 2.2x10^{-16}$ vs $P = 2.8x10^{-14}$, respectively) (Figure 5.18). To further investigate the difference in expression levels, a Wilcoxon signed – rank test was performed on the Ki67 %PI scores at each disease site grouped by the TMA core location. While no statistically significant difference was seen between left and right – sided disease at the invasive margin ($P = 0.4$), right – sided disease showed a trend towards a significantly increased proliferative rate within the tumour core ($P = 0.082$, Figure 5.18).

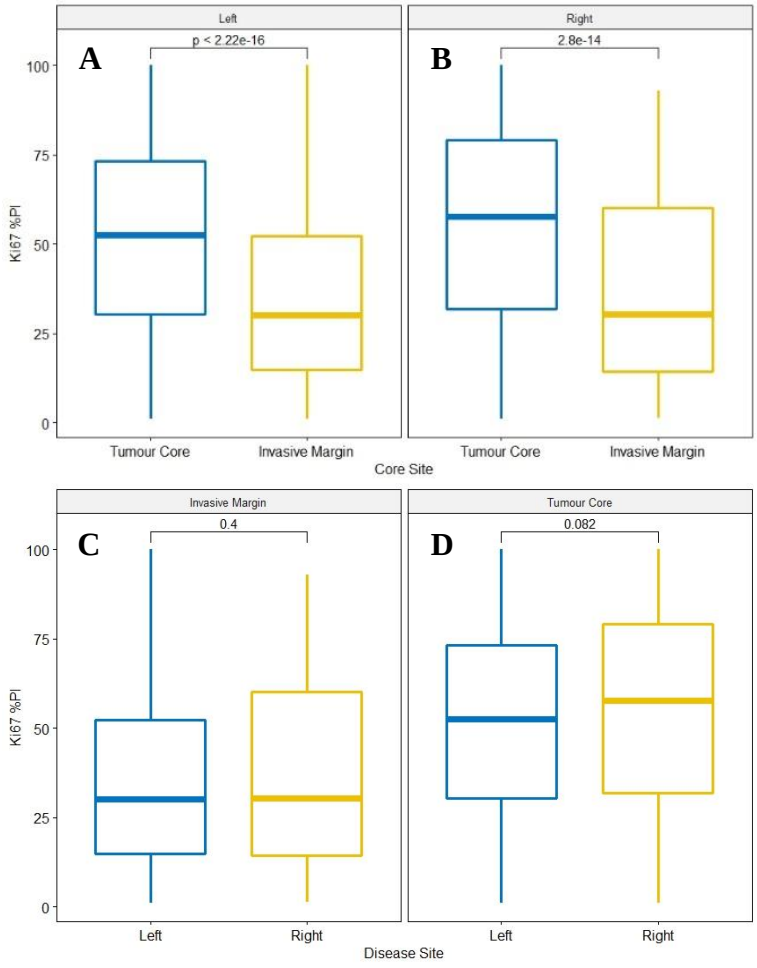


Figure 5.18 Ki67 Expression at the Invasive Margin and Tumour Core in the TransSCOT Cohort. *AI determined Ki67 expression in the IM and TC in left – sided (A) and right – sided disease (B). Ki67 expression in right and left – sided disease at the IM (C) and TC (D). P values are Wilcoxon Matched Pairs Signed – Rank Test (AB) and Wilcoxon Signed – Rank Test (CD).*

Kaplan – Meier survival analysis was used to determine whether site dependent variation in Ki67 expression affected the prognostic significance of the assessment. Ki67 expression retained a statistically significant association with DFS at both the tumour core and invasive margin in left – sided disease with a slight increase in hazard ratio at the tumour core, $HR_{TC} = 1.241$ vs $HR_{IM} = 1.233$, although the association with DFS was more significant at the invasive margin, $P_{TC} = 0.022$ vs $P_{IM} = 0.018$ (Figure 5.19). In right – sided disease however, Ki67 expression did not show a significant association with DFS at either the tumour core or the invasive margin and both locations showed and identical hazard ratio, HR = 1.138 (Figure 5.20).

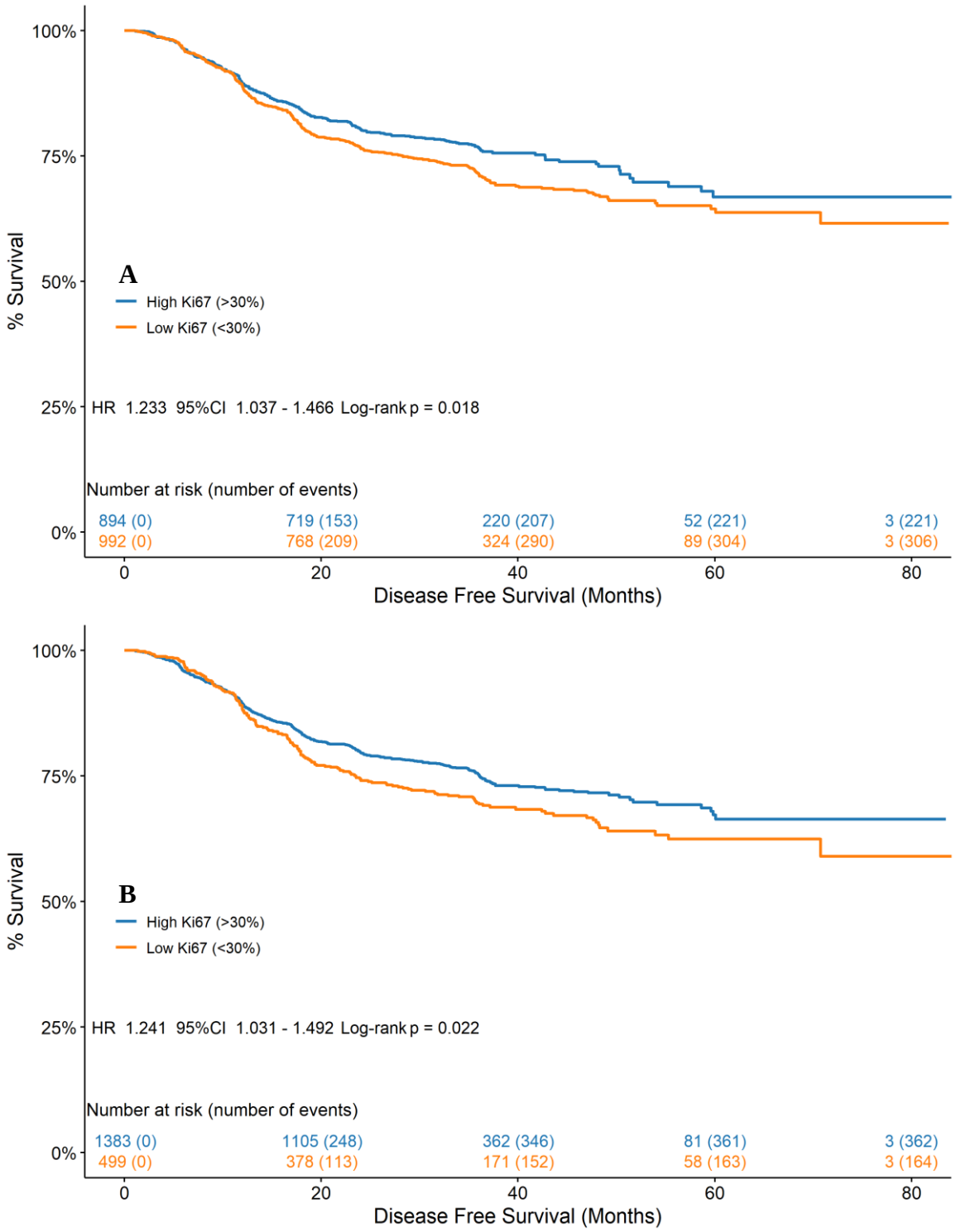


Figure 5.19 Relationship Between Ki67 Expression and DFS in Left – Sided Disease in the TransSCOT Cohort. *Association between AI – determined Ki67 expression and DFS at the invasive margin (A) and tumour core (B) in left – sided disease in the TransSCOT clinical trial cohort.*

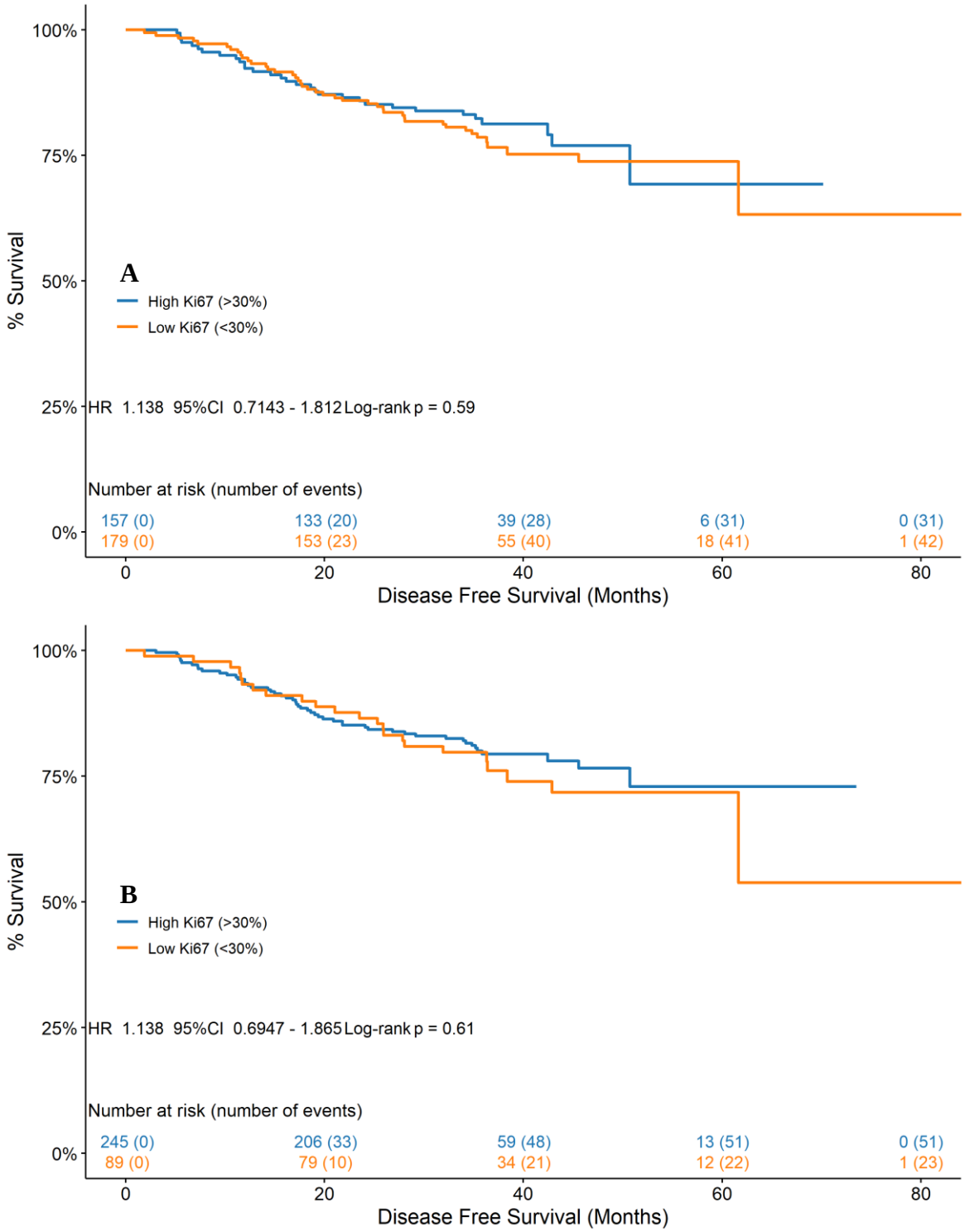


Figure 5.20 Relationship Between Ki67 Expression and DFS in Right – Sided Disease in the TransSCOT Cohort. *Association between AI – determined Ki67 expression and DFS at the invasive margin (A) and tumour core (B) in right – sided disease in the TransSCOT clinical trial cohort.*

5.4 Discussion

Molecular markers of disease progression and response to treatment form the basis of personalised medicine, by providing ever greater insight to the individual's specific condition. Currently, only a few molecular biomarkers are employed in routine pathology of CRC, mainly to assess MMR status through immunohistochemical staining of the four mismatch repair proteins. Recent research has led to the development of a multitude of biomarkers that hold both prognostic and theragnostic significance that are readily translatable to the diagnostic process, notably the proliferation associated antigen Ki67. Given that sustained replicative potential is widely accepted as a hallmark of cancer (Hanahan & Weinberg, 2011), Ki67 aided assessment of the proliferative activity of neoplastic cell populations holds significant clinical relevance. In the setting of breast cancer, Ki67 labelling index is utilised clinically to distinguish between Luminal A and Luminal B intrinsic subtypes and can inform response to endocrine therapy, however, it is noteworthy that recommendations for using Ki67 for this purpose are predicated on good local control of staining quality and to be used diligently due to noted interobserver variability (Goldhirsch et al., 2011).

Image analysis, while still beholden to histological processing quality control, could address the issue of interobserver variability and indeed studies have shown IA algorithms improve the accuracy and reproducibility of Ki67 assessment (Z. Li et al., 2022). Additionally, the time burden associated with Ki67 assessment across WSI renders it clinically impractical leading to the suggestion of conducting the assessment within multiple ROI, the selection of which is subjective in and of itself and has the potential to influence the outcome of the Ki67 indexing (Christgen et al., 2015). This too could be addressed through image analysis to provide an unbiased view of the proliferative activity of the tumour as a whole. In the present study, image analysis tissue and cell segmentation algorithms were utilised to develop a reproducible, automated assessment of the Ki67 proliferative index and assess its efficacy in multiple, distinct CRC patient cohorts.

As stated previously in this work, the use of deep learning on WSI necessitates considerable investment in IT systems that are not feasible everywhere and the benefit – cost trade off must be considered. Therefore, an initial comparative study between a traditional threshold – based cell segmentation algorithm and a deep learning – based cell segmentation algorithm was conducted to assess which approach performed the best in terms of both survival outcomes and statistical agreement with manual histopathological assessment, and if any improvement with deep learning was significant enough to justify its use. When both

algorithms were applied to the TMA constructed from the Glasgow Development Cohort, the Ki67 %PI generated by the deep learning algorithm showed greater agreement with manual histopathological assessment than the traditional algorithm (Figure 5.4) and showed greater association with CSS (Figure 5.8), however, both algorithms failed to outperform manual histopathological assessment for survival stratification (Figure 5.7 & 5.8). Furthermore, optimal cut off analysis of both algorithms showed near – perfect concordance of the deep learning algorithm's cut off value (29.57%) with the 30% utilised in manual histopathological assessment, but notable deviation of the threshold – based algorithm with a value of 18.65% (Figure 5.6), likely a result of less accurate cell segmentation and classification skewing patient Ki67 positivity indices (Bankhead, 2022).

These data were confirmed in the Glasgow Validation cohort, while both algorithms showed excellent and almost identical association with manual histopathological assessment via receiver operator characteristic analysis (Figure 5.9), the AI – based algorithm demonstrated almost perfect statistical concordance with manual histopathological assessment (ICC3 = 0.83) whereas the threshold algorithm demonstrated lower, albeit still substantial, statistical concordance (ICC3 = 0.74). Optimal cut off analysis for CSS again demonstrated the improved concordance of the AI – algorithm with manual histopathological assessment, generating a cut off of 34.51%, while the threshold algorithm showed large discrepancy with the accepted 30% cut off, generating a cut off of 59.37% (Figure 5.11). This discrepancy carried over into Kaplan – Meier survival analysis where the threshold algorithm failed to reach statistical significance for CSS stratification (Figure 5.12), whereas the AI – based algorithm outperformed manual histopathological assessment for patient survival (Figure 5.13). The fact that both algorithms followed an identical workflow until the application of the cell segmentation step likely indicates that the difference in performance is due to the ability of the algorithms to handle the morphological heterogeneity of cell nuclei both within the same cohort, for example high vs low grade dysplasia, and across independent cohorts. Both Glasgow cohort TMAs were of similar age, processed similarly, and of comparable scanning quality, however, the consistent performance of the deep learning algorithm between the cohorts demonstrates that this approach is able to handle histological variation better than traditional threshold – based approaches.

As previously emphasized in this work, translatability of image analysis algorithms is vital to their use in research, as companion diagnostics in clinical trials, and indeed in aiding primary diagnosis. Reliance on manual histopathological assessment being conducted prior to image analysis poses a two – fold barrier to their continued development, in terms

of the time taken to produce a second score for the same patients and limiting the validity the performance of image analysis algorithms for survival prediction if statistical measures of agreement are subjectively considered suboptimal. Having validated the AI algorithm in two clinical cohorts, it was subsequently applied to two more clinical cohorts, one of which was previously unassessed for Ki67 and one where Ki67 assessment had been conducted on full sections but not on the TMA used in this study. In the previously unscored GRI cohort, the AI algorithm generated Ki67 %PI scores showed a statistically significant stratification for CSS using the 30% cut off value utilised in previously cohorts (Figure 5.15). Additionally, the novel cut off value generated for this cohort from the AI generated scores was nearly identical (30.4%) to the value manually determined in previous cohorts (Figure 5.14), further demonstrating that image analysis algorithms can reliably perform primary research for novel cohorts and data.

Further validation in the TransSCOT clinical trial cohort demonstrated the AI algorithm's ability to prognostically stratify patients on novel image data in a cohort where prior confirmation of Ki67 expressions association with DFS was performed on a subset of WSI. While the optimal cut off values generated for this cohort were notably lower than the 30% PI determined in prior cohorts (Figure 5.16), the algorithm showed prognostically significant stratification using the 30% cut off at both the tumour core and invasive margin (Figure 5.17). This discrepancy between the generated cut offs and the accepted 30% value is likely due to the tendency of hazard ratios to increase with a large difference of patient numbers between the stratified patient groups. Closer inspection of the maximally ranked distribution of events within the two tumour subsites shows that the second highest association of events with Ki67 expression occurs closer to the 30% value (Figure 5.16).

Previous studies have demonstrated variation in Ki67 expression between left and right – sided disease, tumour subsite, and within the crypts of pre – carcinomatous polyps, although with conflicting results (Davenport et al., 2003; De Jong et al., 1998; Gao et al., 2017; Nayak et al., 2021). Investigation of this phenomenon in the TransSCOT cohort revealed that Ki67 expression analysis showed a marginal increase in prognostic significance when conducted at the invasive margin over the tumour core (Figure 5.17). Given that this study was conducted on a single TMA core taken at both tumour subsites for each patient, it would be of interest to further investigate this effect using WSI to establish a more complete understanding of the potential difference in prognostic significance, as the majority of current Ki67 assessment is conducted within the tumour core. Concurring with findings that Ki67 %PI values are significantly lower at the invasive margin (Brabletz et al., 2001; A.

Jung et al., 2001), evaluation of localization dependent Ki67 expression demonstrated a significantly higher mean Ki67 %PI in the tumour core in both left and right – sided disease (Figure 5.18 A, B), with the difference more pronounced in left – sided disease though this is likely due to the higher numbers of patients with left – sided disease. Analysis of Ki67 expression localization by tumour site revealed no significant difference in %PI values between right and left – sided disease at the invasive margin however, Ki67 %PI values in the tumour core showed a trend towards significantly higher expression in right – sided disease. While some studies have noted no difference in site dependent Ki67 expression (Shin et al., 2014), others have demonstrated significantly increased proliferative activity in MSI tumours which are predominantly right – sided (Michael-Robinson et al., 2001; Takagi et al., 2002). Unfortunately, microsatellite stability status was not available for these patients, therefore it would be of interest in future work to investigate whether this effect is microsatellite stability or tumour site dependent and if this has a bearing on survival.

Kaplan – Meier survival analysis of tumour subsite dependent Ki67 expression revealed an inverse pattern of correlation with DFS in left – sided disease than a combination of both disease sites, whereby Ki67 expression assessed within the tumour core demonstrated a greater hazard ratio than that of the invasive margin, however, the association was marginally statistically more significant at the invasive margin (Figure 5.19). In right – sided disease, neither analysis of Ki67 expression at the tumour core or invasive margin showed a statistically significant association with DFS and indeed produced identical hazard ratios (Figure 5.20). While this is likely due to the comparatively low number of patients with right – sided disease (N = 336, 11.53% of TransSCOT cohort), it does not negate the possibility that tumour site and subsite specific determination of optimal cut off values is required to adequately assess the association of Ki67 expression in these patients, given that optimal cut off values have been previously been determined in large, non – subsetted clinical cohorts that are skewed towards left – sided disease.

A limitation of the present study that was unable to be addressed in the algorithm development process and one that could potentially improve the validity of this study and those like it is the appropriate determination of a DAB – colour deconvolution threshold value for positive cell classification. The threshold selected in this algorithm was determined by applying the work in progress algorithm to multiple TMA cores from the Glasgow Development cohort that exhibited variation in stain intensity and assigning a value that correctly classified the weakest Ki67 stained cell as positive, over and above any background non – specific staining levels. While the researcher has performed prior manual

histopathological assessment of Ki67 expression, this method of threshold selection is still conspicuously subjective in nature and was performed only in a single cohort, where section age, storage, and histological processing could affect the quality of staining and thus the prognostic outcome of the assessment. Therefore, to optimize the selection of a DAB threshold, multiple observers could be utilized to classify cells as positive or negative, across various cohorts that have undergone identical histological processing, to produce a consensus value for Ki67 positivity that would go some way to further reduce the subjectivity of the image analysis.

In summary, this chapter demonstrates the reliability and reproducibility of deep learning – based image analysis for Ki67 expression assessment across multiple cohorts. Data from the Glasgow Development and Validation cohorts shows that image analysis is able to outperform manual histopathological assessment for prognostic patient stratification while maintaining excellent measures of statistical agreement. Applying the deep learning – based algorithm to the previously unscored GRI cohort confirms the ability of the algorithm to conduct primary research of novel image data and patient cohorts in concordance with expected outcomes determined in previous cohorts, which was further confirmed in the TransSCOT clinical trial cohort. In addition to prognostically stratifying patients for DFS, data from the TransSCOT cohort highlighted possible differences in Ki67 expression between tumour site and localization within the tumour that could provide more targeted risk stratification methods based on established prognostic criteria utilized in routine diagnosis. As discussed above, future work should be conducted to optimize the cell classification step of this algorithm to provide an increasingly less subjective method of Ki67 IHC assessment, potentially serving as a framework for developing and validating algorithms targeted towards other molecular markers of prognosis.

6. Image Analysis Approaches to Histological CRC Prognostic Classification Systems

6.1 Introduction

The Tumour Node Metastasis (TNM) staging criteria is the gold standard for CRC prognosis and direction of therapeutic decisions, utilised in diagnostic pathology throughout the world. However, issues in determining prognosis and appropriate therapy regimens for Stage II disease has long presented an unmet need for additional criteria that is able to adequately direct care for these patients. An early study to address this was conducted by Petersen and colleagues who identified a combination of pathological features that were able to identify high – risk Stage II patients who could benefit from adjuvant chemotherapy (Petersen et al., 2002). This system assigned a point each to peritoneal involvement, submucosal and extramural venous invasion, marginal involvement, and two additional points if tumour perforation of the bowel was present, resulting in a five – point score that was dichotomized at ≥ 2 points for risk stratification.

Since this study, the continued development of omics techniques for interrogating tumours in ever finer detail has steered research towards developing similar subtyping methodologies based on patients' gene expression profiles. In 2015, the CRC Subtyping Consortium (CRCSC) performed network clustering on six independent, gene expression – based subtyping algorithms applied to over 3000 patients across multiple platforms and sample types, to study the association of the respective subtypes (Guinney et al., 2015). The result of this work was the four Consensus Molecular Subtypes (CMS) of CRC denoted; MSI Immune (CMS1), Canonical (CMS2), Metabolic (CMS3), and Mesenchymal (CMS4). The four subtypes were determined based on shared mutational characteristics but were subsequently found to share additional prognostic, theragnostic, and phenotypic attributes. While this subtyping approach demonstrated the importance of precision medicine methodologies in the continued development of CRC research, its costly and time – consuming implementation coupled with the issue that ~13% of patients remain unclassified as neither outliers nor a fifth subtype, has made implementation of the CMS in routine diagnostics unfeasible.

In 2017, Isella and colleagues postulated that the use of whole tumour lysates in the datasets used to generate the CMS introduced a large number of genes of stromal origin in CMS4 likely masking the gene expression profiles of the tumour cells themselves (Isella et al., 2017b). To ameliorate the effect of stroma associated genes and generate a more tumour specific means of CRC subtyping, the transcriptomes of 515 samples (from 244 patients) of patient derived xenografts (PDXs), wherein the stromal component of the tumour is replaced with murine constituents as part of the xenotransplantation process, were analysed using

human – specific probes and a non – negative matrix factorization clustering algorithm to assign each sample to one of five classes. The result of this work was the CRC Intrinsic Subtypes (CRIS) denoted A – E, each of which are enriched for expression of genes / mutations conveying distinct phenotypes and prognoses in addition to carrying a greater classification rate than the CMS (94% vs 87%, respectively). While the use of PDXs to reduce the influence of stroma originated genes likely provides a more accurate interpretation of the mutational landscape of the tumour than the CMS, it goes no further to providing a clinically actionable subtyping methodology. The CRIS is still hamstrung by the use of expensive transcriptomic analyses and indeed the authors additionally acknowledge that the xenotransplantation process could alter and drive cellular processes that form some constitutive features of specific subtypes.

Studies looking to establish histology – based subtyping systems that provide equal prognostic information as the CMS / CRIS but would inherently be more readily translatable to routine diagnostics have yielded two notable methodologies. Park and colleagues in 2015 first described the Glasgow Microenvironment Score (GMS), a combinatorial score of two well established prognostic factors assessed on diagnostic H&E sections, the Tumour Stroma Percentage (TSP) and the Klintrup – Mäkinen (KM) grading system of peritumoural inflammatory response (Park et al., 2015b). This system significantly stratifies patients for prognosis into three distinct groups; patients with a high KM grade are assigned GMS 0 conferring a good prognosis, patients with a low KM grade but high TSP are assigned GMS 2 conferring the worst prognosis, and patients with low scores for both metrics are assigned GMS 1 with intermediate prognosis. In addition to prognosis, this system has recently been shown to hold theragnostic significance, with the inflammation dense GMS 0 possessing a significantly better prognosis when administered FOLFOX chemotherapy over CAPOX (Alexander et al., 2021a).

In a study reviewing the CMS, aiming to translate the phenotypic features of each subtype to histological assessment, Roseweir and colleagues identified prominent immune infiltration and stromal density as concurrent characteristics of CMS 1 / 4 and GMS 0 / 2 and additionally noted that increased proliferative activity was present in CMS 2. Given that other previous methodologies have identified subtypes aligning with CMS 2 that exhibit increased proliferation demonstrable through IHC – based Ki67 positivity indexing, this marker was used to separate GMS 1 into two prognostically distinct subtypes (Roepman et al., 2014b). This work yielded the four Phenotypic Subtypes of CRC, denoted Immune (high KM grade and best prognosis), Canonical (high Ki67 index and good prognosis), Latent (low

for all measures and poor prognosis), and Stromal (high TSP and worst prognosis). This subtyping methodology is independently prognostic, able to predict the risk of recurrence, and the Immune subtype still exhibits the same response to chemotherapeutics as GMS 0, thus offering significant clinical information about the TME in addition to being extremely translatable to routine diagnostic pathology.

The aim of the present chapter, therefore, is to amalgamate the algorithms and data described and generated in previous chapters to digitally interpret these histology – based subtyping systems through image analysis, to compare their prognostic and theragnostic capacity to that of manual histopathological assessment, and to evaluate the clinical efficacy and applicability of different automation methods of the digital systems.

6.2 Materials and Methods

6.2.1 Histopathological Assessment

Manual Histopathological assessment of the TSP and KM grade was conducted on H&E WSI for all cohorts as previously described in 3.2.1 and 4.2.1, respectively. For the TSP briefly, on the slide showing the deepest point of tumour invasion, a representative intratumoural area was selected and an FOV with tumour cells at all four edges at 10x magnification was assessed for the percent (to the nearest 5%) of desmoplastic stroma occupying the field. A cut off of </> 50% stromal content was used to stratify patients for high vs low stromal component. For the KM grading, the invasive margin on the slide showing the deepest point of tumour invasion was assessed for lymphocyte response using a four point scale as follows: 0 – no evidence of inflammatory response, 1 – patchy distribution of lymphocytes across the invasive margin, 2 – notable inflammatory response forming a band – like structure across a significant portion of the invasive margin, 3 – prominent inflammatory response forming a florid cup – like structure across the invasive margin with possible evidence of tumour island destruction. To stratify patients for high vs low inflammatory response, the four – point scale was amalgamated into "weak" inflammatory response (KM grades 0/1) and "strong" inflammatory response (KM grades 2/3).

IHC staining of all CRC tissue for Ki67 utilised in this study was conducted according to the methodology described in 2.2.2. Histopathological assessment of Ki67 positivity for all cohorts was conducted as described in 2.3.3. Briefly, for TMAs, 100 cells were counted at random and the % positivity was determined for each individual core, where multiple cores were available for a single patient, the % positivity was counted across all cores where sufficient tissue was available and averaged to determine the % positivity for each patient.

6.2.2 Statistical Analysis

Patient cohort clinicopathological characteristics and test of association with CSS / DFS were determined in IBM SPSS version 25 (IBM, New York, USA). P values of <0.05 were considered to demonstrate a statistically significant association between clinicopathological features and clinical outcomes. All other statistical analyses for this study were conducted in RStudio (RStudio, Boston, MA). Survival analysis was conducted and associated Kaplan – Meier survival curves produced using the survival, survminer, and survMisc packages, with the log rank statistic used to assess association between classification system stratification and CSS / DFS. Hazard ratios and associated 95%

confidence intervals for survival analysis were calculated using the Cox proportional hazard model.

6.3 Results

6.3.1 Image Analysis Approaches to the Glasgow Microenvironment Score in the Glasgow Development Cohort

As described in 6.1, research into the development of novel prognostic systems for CRC has yielded promising results through both classical histology – based assessment and modern omics – based analyses, however, both have issues prohibiting their imminent translation to routine diagnostics. Omics – based analyses are extremely targeted, and in some cases provide an arguably more actionable overview of the nature of an individual's tumour, but the methodologies used to determine patient grouping are at present too costly to implement in pathology labs. Histology – based systems for the most part do not require any additional assays or tissue to stratify patients, indeed in the cases where additional investigation is necessary, the only requirement is additional IHC sections from the diagnostic block, which is already part of routine diagnosis. Furthermore, the time required of the pathologist to conduct additional assessment beyond the AJCC staging criteria must be a consideration for implementing such novel classification systems. Therefore, in the present chapter, two classification systems determined via two approaches – one semi – automated requiring an invasive margin annotation and fully automated requiring no human input prior to analysis – were assessed for prognostic ability.

The Glasgow Microenvironment Score (GMS) is a combinatorial score of the TSP and KM assessments conducted on a single H&E slide / WSI that stratifies patients into one of three prognostic groups. The assessment was developed by Park and colleagues in 2015 (Park et al., 2015a), and has subsequently been independently validated in multiple retrospective clinical and clinical trial patient cohorts (Alexander et al., 2021a). Scoring is conducted by initially assessing patients for the KM grade and all patients with a high inflammatory response, regardless of TSP status, are assigned GMS 0. The TSP of the remaining patients is then used to assign the remaining two groups, with patients with a high TSP (>50%) assigned GMS 2 and patients with low scores for both metrics assigned GMS 1. This method and associated prognoses are detailed in Table 6.1.

	GMS 0	***GMS 1***	***GMS 2***
KM Grade 0/1 vs 2/3	High	Low	Low
TSP Status </> 50%	Any	Low	High
Cancer Specific Prognosis	Good	Intermediate	Poor

Table 6.1 Overview of Glasgow Microenvironment Score. *Components and scoring method of the Glasgow Microenvironment Score and associated cancer specific prognoses.*

Two methodologies were utilised to assess the GMS via image analysis, both utilising a single H&E WSI, and both applying the deep learning – based TSP algorithm (detailed in 3.3.3) to categorize patients as high and low for relative stromal component. The first methodology took a semi – automated approach to patient categorization for peritumoural inflammatory response, requiring annotation of the invasive margin to generate an ROI within which the lymphocyte detection algorithm was applied, following an identical workflow to that described in 4.3.1. The same cut off value was used to stratify patients into high vs low immune infiltrate as in 4.3.1 and the TSP algorithm was used to stratify patients as described in 3.3.3. These two stratified metrics were then used to generate a GMS group for each patient. The overall GMS algorithm workflow is detailed in Figure 6.1.

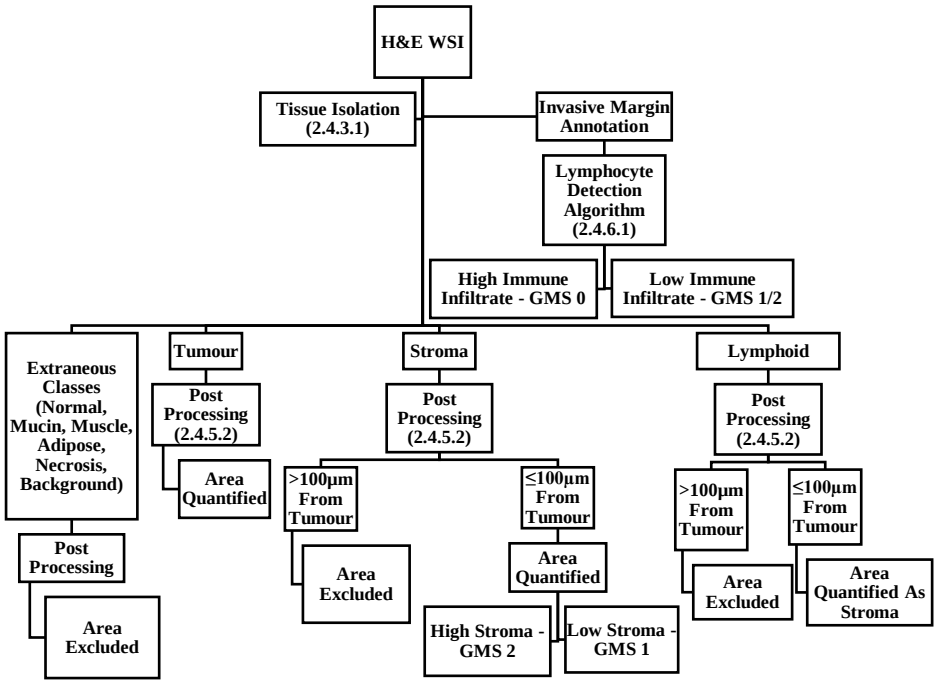


Figure 6.1 Semi – Automated GMS Algorithm Workflow. *Workflow of semi – automated image analysis GMS algorithm. Algorithm is a combination of deep learning TSP algorithm (3.3.3) and lymphocyte detection algorithm conducted in invasive margin annotations (4.3.1).*

The primary outcome of all classification system studies was association with CSS / DFS and comparison thereof to survival outcomes determined by manual histopathological assessment. Therefore, Kaplan – Meier survival analysis was conducted on patients from the Glasgow Development Cohort that had been categorized using the GMS image analysis algorithm described above to determine its association with CSS. The GMS algorithm stratification showed a statistically significant association with GMS and additionally, demonstrated a greater hazard ratio than manual histopathological assessment for stratification (HR 2.031 95% CI 1.472-2.801 $P < 0.0001$ vs HR 1.901 95% CI 1.403-2.576 $P < 0.0001$, respectively) (Figure 6.2).

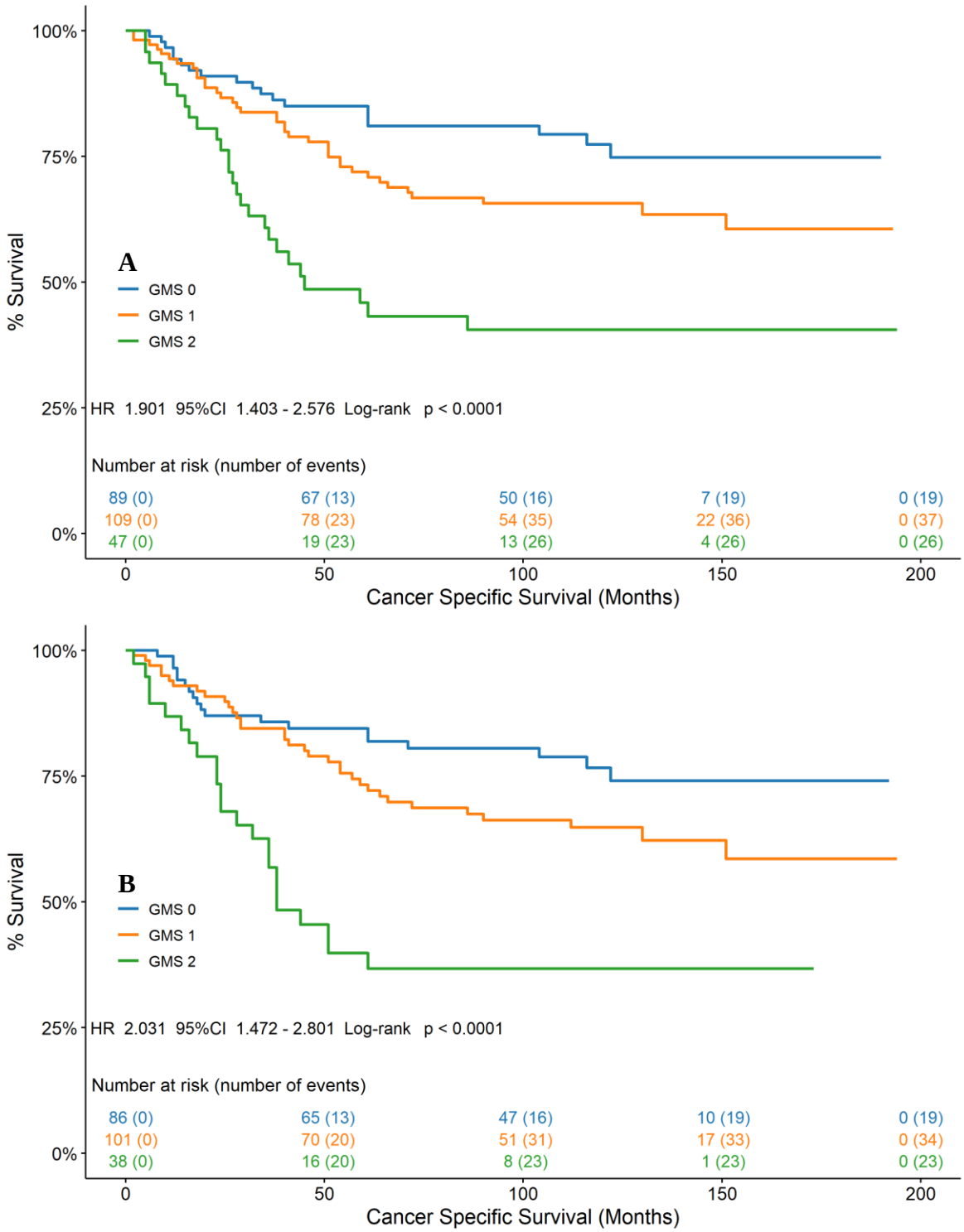


Figure 6.2 Relationship Between CSS and GMS in Glasgow Development Cohort. *Association of CSS and GMS determined via manual histopathological assessment (A) and semi – automated image analysis assessment (B) in the Glasgow Development Cohort.*

The second methodology used to stratify patients for the GMS criteria using image analysis took a fully automated approach to determine the GMS group from a single H&E WSI. This workflow again utilised the deep learning TSP algorithm coupled with the automated lymphocyte detection algorithm described in 4.3.4. To conduct the GMS assessment using this methodology, the deep learning TSP algorithm was initially applied to the WSI, and the stroma class overlays generated by the algorithm were converted to ROI for analysis, within which the same lymphocyte detection algorithm (4.3.1) was applied. The overall workflow for this algorithm is detailed in Figure 6.3.

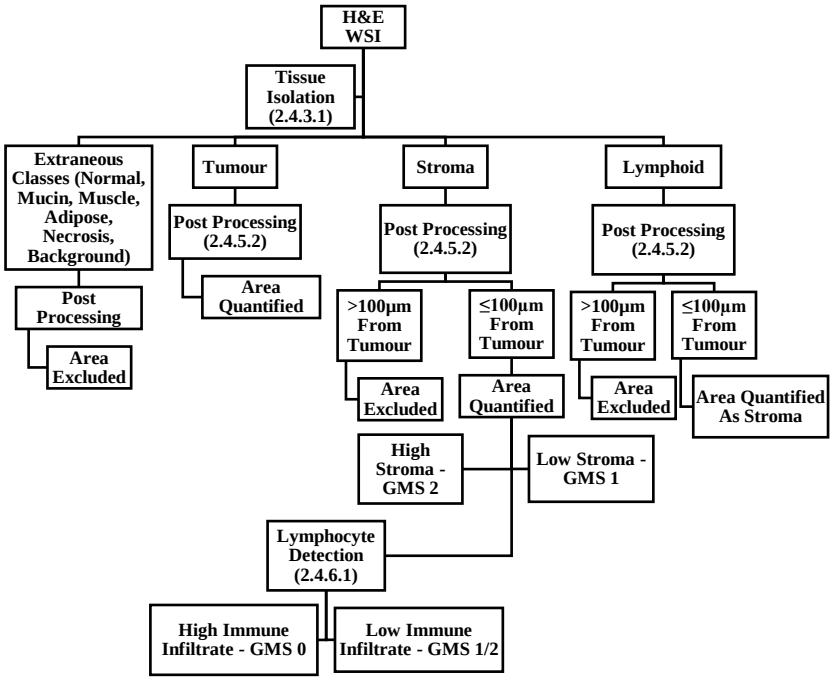


Figure 6.3 Automated GMS Algorithm Workflow. *Workflow of automated GMS image analysis algorithm coupling deep learning TSP and TIL assessment. Methods sections detailing operations are given in brackets.*

Kaplan – Meier survival analysis was used to assess the association of the stratification produced by the automated GMS algorithm and CSS. The automated image analysis demonstrated a statistically significant association with CSS however, this approach did not outperform the manual histopathological assessment (HR 1.651 95% CI 1.215-2.243 P = 0.00033 vs HR 1.901 95% CI 1.403-2.576 P < 0.0001, respectively) (Figure 6.4).

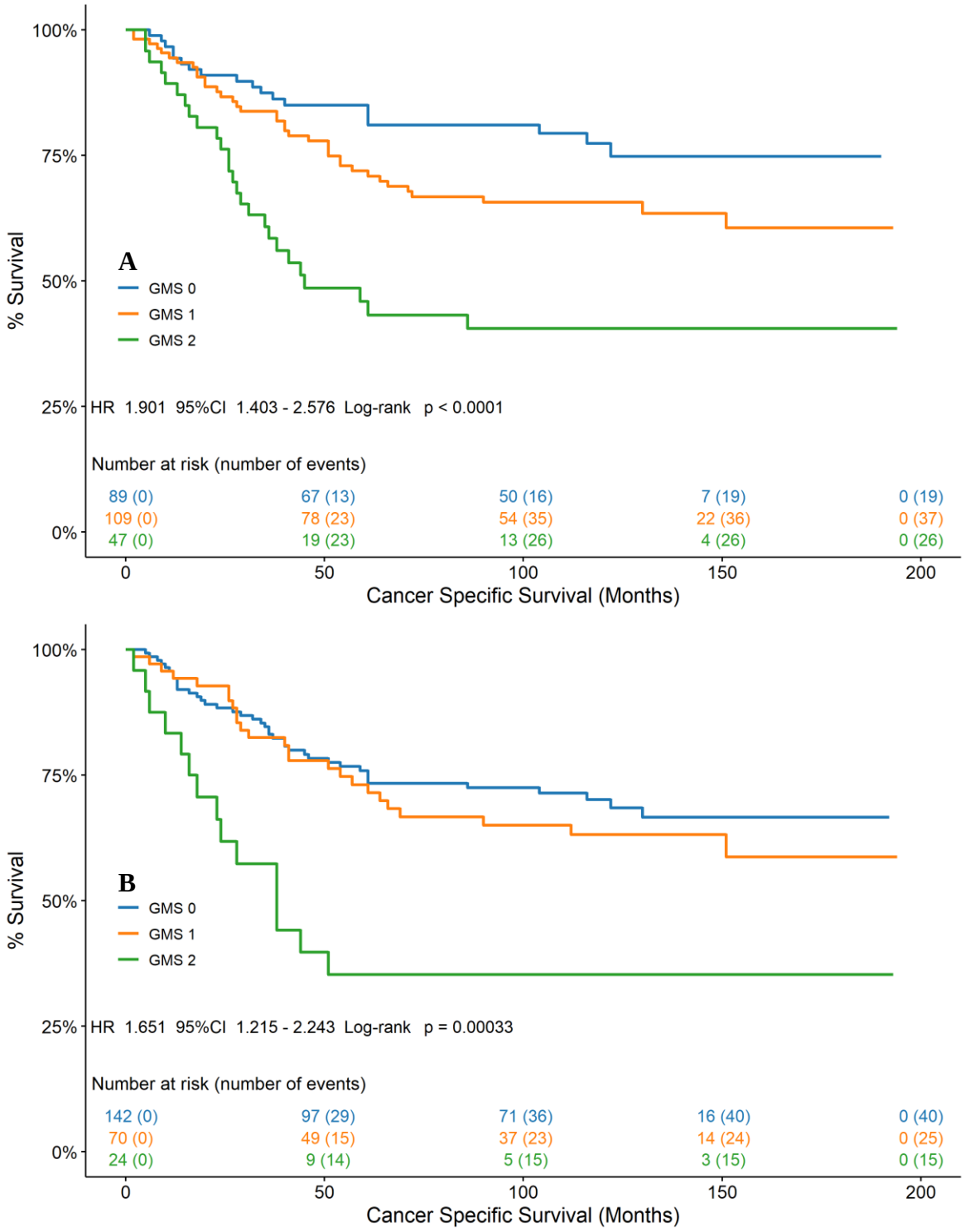


Figure 6.4 Relationship Between CSS and GMS in Glasgow Development Cohort. *Association of CSS and GMS determined via manual histopathological assessment (A) and fully automated image analysis assessment (B) in the Glasgow Development Cohort.*

6.3.2 Validation of Image Analysis GMS Assessment in Glasgow Validation Cohort

Having demonstrated that both the semi and fully automated image analysis approaches to the GMS criteria show statistically significant associations with CSS in the Glasgow Development Cohort, the same approaches were then assessed for prognostic ability in the Glasgow Validation Cohort. The semi – automated approach followed an identical workflow to that implemented in 6.3.1 on the same sections identified as appropriate for image analysis in 4.3.2. Kaplan – Meier survival analysis of the GMS stratification determined by image analysis showed a statistically significant association with CSS, but did not outperform the manual histopathological assessment (HR 1.365 95% CI 1.089-1.711 P = 0.016 vs HR 1.928 05% CI 1.484-2.505 P<0.0001, respectively) (Figure 6.5).

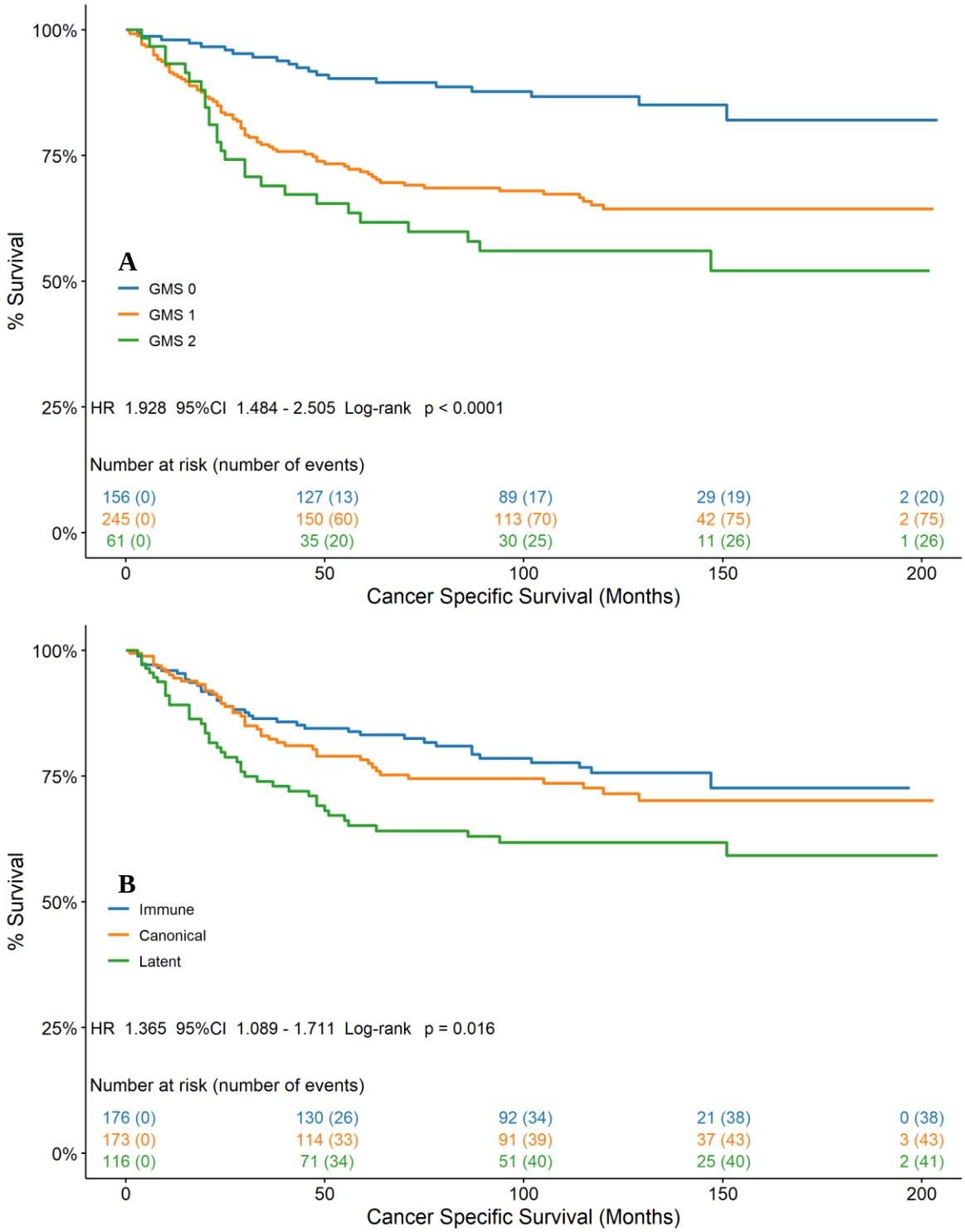


Figure 6.5 Relationship Between CSS and GMS in Glasgow Validation Cohort. *Association of CSS and GMS determined via manual histopathological assessment (A) and semi – automated image analysis assessment (B) in the Glasgow Validation Cohort.*

The fully automated GMS workflow detailed in 6.3.1 was applied to the same Glasgow Validation cohort sections as the semi – automated algorithm, identified in 4.3.2. Kaplan – Meier survival analysis of the GMS stratification produced by fully automated analysis again showed a statistically significant association with CSS, but again did not outperform the manual histopathological assessment (HR 1.31 95% CI 1.053-1.631 P =

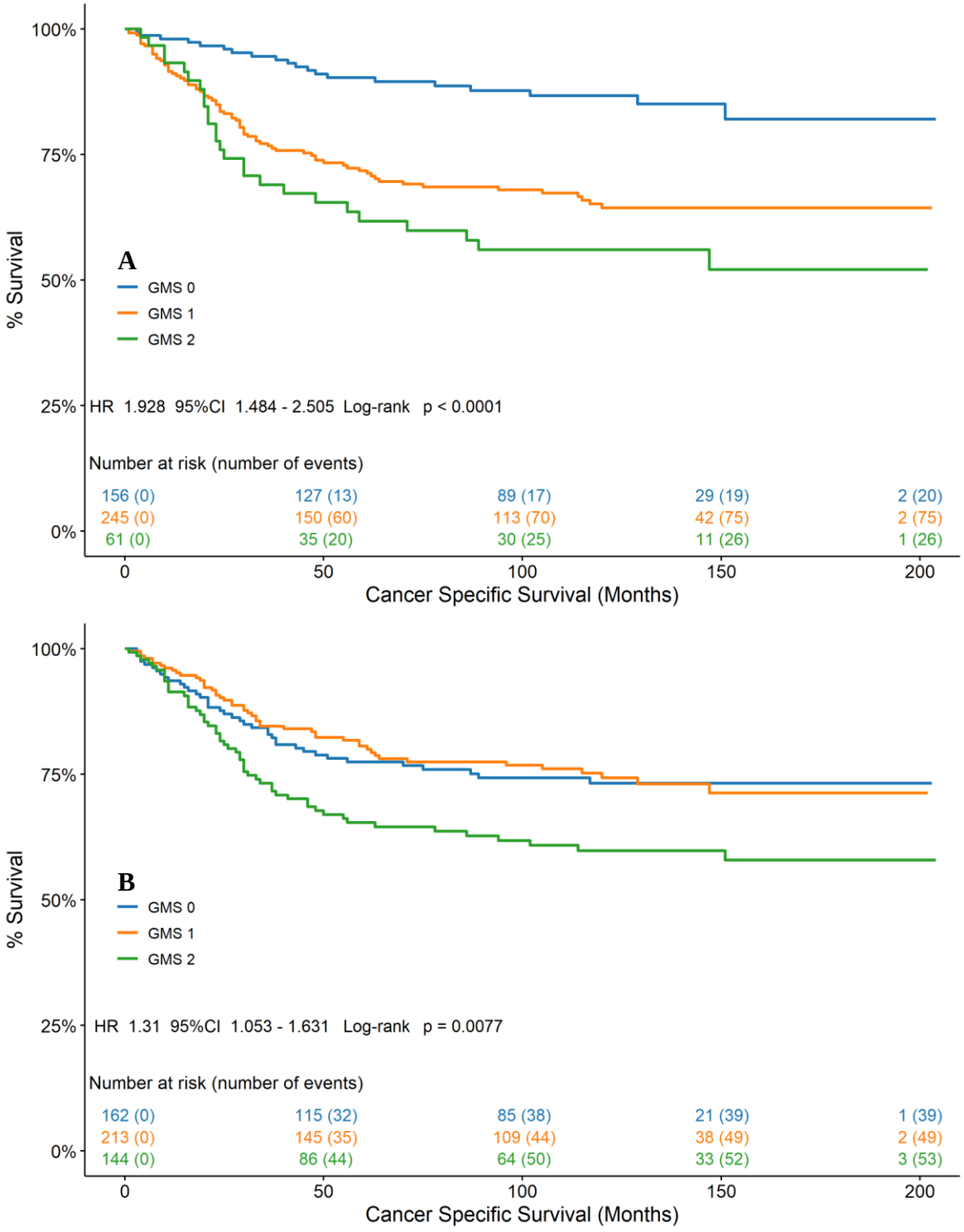


Figure 6.6 Relationship Between CSS and GMS in Glasgow Validation Cohort. *Association of CSS and GMS determined via manual histopathological assessment (A) and fully automated image analysis assessment (B) in the Glasgow Validation Cohort.*

6.3.3 Translation of Image Analysis GMS Assessment to TransSCOT Clinical Trial Cohort

Final validation of the GMS algorithms was performed in the TransSCOT clinical trial cohort, comprising substantially more patients than the Glasgow Development and Validation cohorts. Although both the TSP and lymphocyte detection algorithms that constitute the GMS methodology have been previously validated in this cohort, it is difficult to predict the interaction of the two algorithms when stratifying for prognosis in a multi – stage system applied to large patient numbers. The translational ability of image analysis algorithms of this type is contingent on both the constituent components being individually prognostic and retaining the prognostic capacity when used in combination. To validate the GMS algorithm in this cohort, the semi and fully automated methods were applied to the H&E WSI using an identical workflow to that described in 6.3.1 and 6.3.2.

Initially, the semi – automated algorithm was used to stratify patients into the GMS groups and analysed for survival prediction. Kaplan – Meier survival analysis of the GMS stratification generated by this algorithm showed a statistically significant association with DFS, performing comparably to manual histopathological assessment (HR 1.475 95% CI 1.289-1.688 $P < 0.0001$ vs HR 1.494 95% CI 1.314-1.699 $P < 0.0001$, respectively) (Figure 6.7).

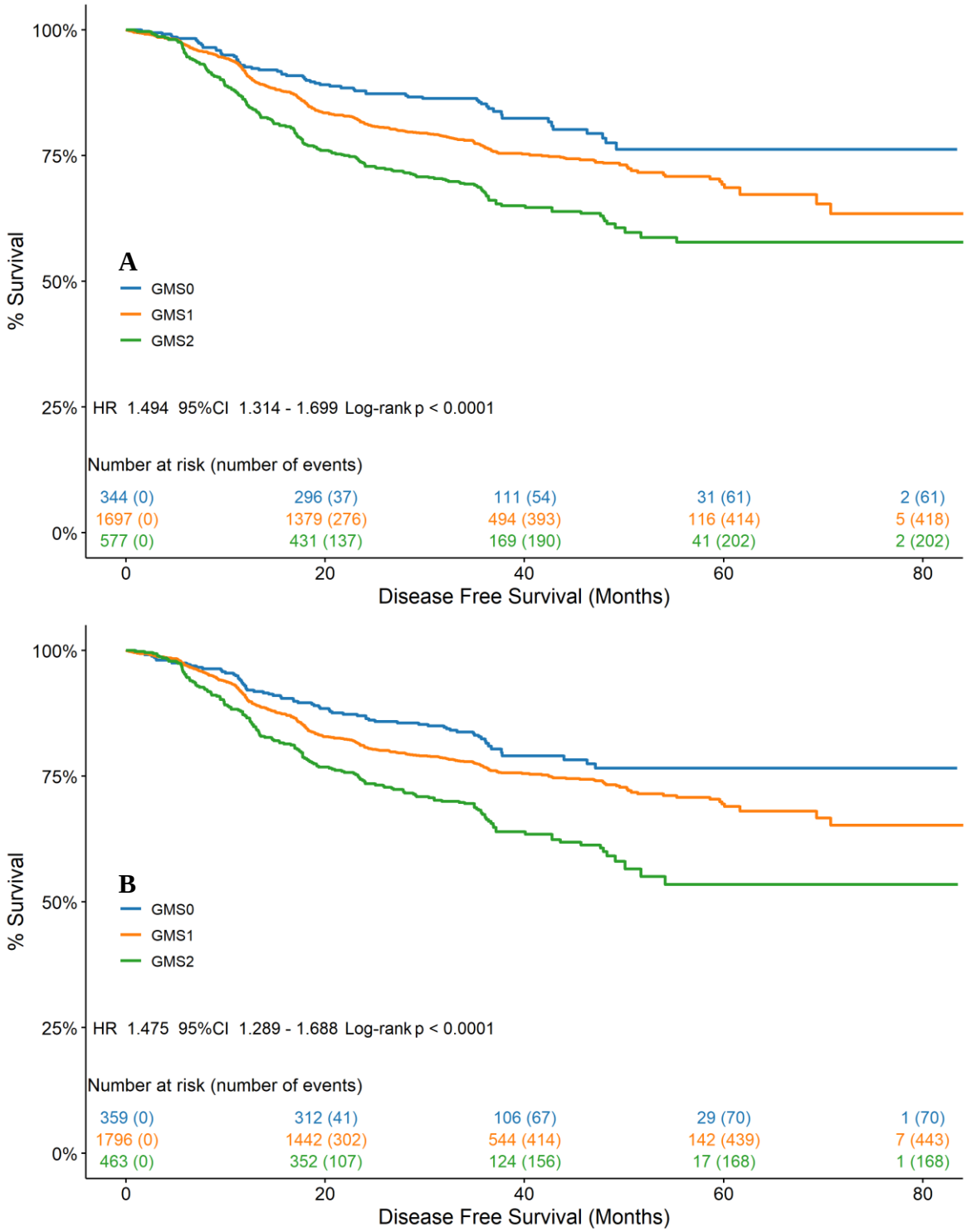


Figure 6.7 Relationship Between DFS and GMS in TransSCOT Clinical Trial Cohort. *Association of DFS and GMS determined via manual histopathological assessment (A) and semi – automated image analysis assessment (B) in the TransSCOT Clinical Trial Cohort.*

The fully automated GMS algorithm was then applied to the same H&E WSI and analysed for DFS prediction. Kaplan – Meier survival analysis of the GMS stratification determined by this algorithm again demonstrated a statistically significant association with DFS but demonstrated a moderately lower hazard ratio than manual histopathological assessment (HR 1.312 95% CI 1.189-1.449 P < 0.0001 vs HR 1.494 95% CI 1.314-1.699 P < 0.0001, respectively) (Figure 6.8).

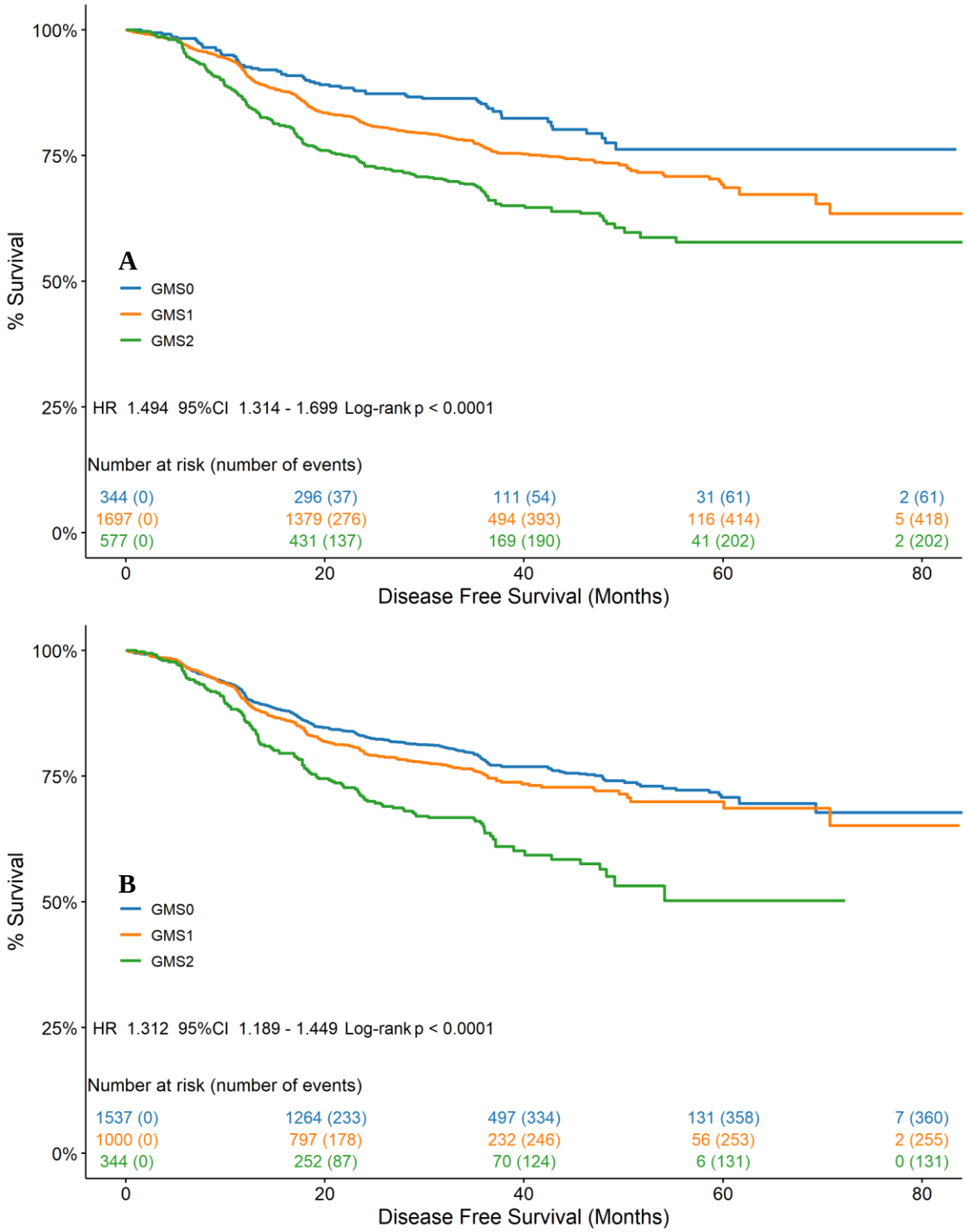


Figure 6.8 Relationship Between DFS and GMS in TransSCOT Clinical Trial Cohort. *Association of DFS and GMS determined via manual histopathological assessment (A) and fully automated image analysis assessment (B) in the TransSCOT Clinical Trial Cohort.*

As it was the most prognostic GMS model, the semi-automated algorithm underwent univariate and multivariate Cox regression analysis to determine the relationship with DFS and prognostic independence from other clinicopathological features. On multivariate analysis, the semi-automated GMS algorithm was found to be significantly associated with DFS (HR 1.722 95%CI 1.335-2.221, *p < 0.001*), independent of T-Stage (*p < 0.001*), N-Stage (*p < 0.001*), and disease site (*p = 0.029*), Table 6.2.

	Univariate				**Multivariate**		
	HR	95% CI	p-value		HR	95% CI	p-value
Sex							
Female	1.0						
Male	1.040	0.898-1.205	0.599				
T Stage				**T Stage**			
1	1.0			1	1.0		
2	1.013	0.498-2.060	0.973	2	1.121	0.551-2.281	0.753
3	1.85	0.988-3.467	0.055	3	1.923	1.025-3.608	**0.042**
4	3.217	1.715-6.037	**<0.001**	4	3.290	1.745-6.205	**<0.001**
N Stage				**N Stage**			
0	1.0			0	1.0		
1	1.436	1.155-1.787	**0.001**	1	1.783	1.426-2.229	**<0.001**
2	2.821	2.249-3.538	**<0.001**	2	3.002	2.386-3.777	**<0.001**
Site				**Site**			
Left	1.0			Left	1.0		
Right	0.687	0.557-0.848	**<0.001**	Right	0.785	0.632-0.975	**0.029**
Treatment							
FOLFOX	1.0						
CAPOX	1.084	0.924-1.272	0.319				
GMS				**GMS**			
0	1.0			0	1.0		
1	1.342	1.070-1.683	**0.011**	1	1.310	1.044-1.644	**0.02**
2	2.137	1.663-2.746	**<0.001**	2	1.722	1.335-2.221	**<0.001**

Table 6.2 Relationship between GMS, Clinicopathological Features and DFS in TransSCOT Cohort. Hazard ratios and 95% confidence intervals determined via Cox proportional hazards regression and multivariate analysis conducted using backwards conditional method.

6.3.4 Image Analysis Approaches to the CRC Phenotypic Subtypes in the Glasgow Development Cohort

A study by Roseweir and colleagues looking to translate the unique phenotypes associated with each CMS group initially identified notable parallels, namely the prominent inflammatory infiltrate and high stromal component, between CMS 1 / GMS 0 and CMS 4 / GMS 2 (Roseweir et al., 2018b). High proliferative activity was noted as a distinct phenotypic feature of CMS 2 and subsequently utilised Ki67, a proliferation marker that has been clinically validated in other tumour types, to differentiate between good and poor prognosis in GMS 1. This work established the Phenotypic Subtypes of CRC, a novel four – group prognostic system that is readily translatable to routine diagnostics. Manual histopathological assessment of the Phenotypic Subtypes initially follows the same steps as the GMS, whereby the H&E is first assessed for peritumoural inflammatory response by the KM grading criteria, with all patients with a high inflammatory infiltrate (KM 2/3) assigned the Immune subtype. The TSP is then used to stratify patients further, with all high TSP (>50% stromal content) assigned the Stromal subtype, and finally Ki67 is used to stratify the remaining patients with high Ki67 %PI (> 50% in the initial study but subsequent studies determined a consensus cut off value of > 30%) patients assigned the Canonical subtype, and the remaining patients with low scores for all metrics assigned the Latent subtype. The grading criteria and associated subtype prognoses are detailed in Table 6.2.

	Phenotypic Subtype			
	Immune	***Canonical***	***Latent***	***Stromal***
KM Grade 0/1 vs 2/3	High	Low	Low	Low
TSP Status </> 50%	Any	Low	Low	High
Ki67 Proliferation Index </> 30%	Any	High	Low	Any
Cancer Specific Prognosis	Best	Good	Poor	Worst

Table 6.2 Overview of CRC Phenotypic Subtypes. *Assessment criteria and associated prognosis of CRC Phenotypic Subtypes established by Roseweir and Colleagues* (Roseweir et al., 2018b).

To determine the Phenotypic Subtypes via image analysis, the initial steps were the same as those used to conduct the GMS assessment, using semi – automated and fully automated algorithms for lymphocyte infiltrate assessment (4.3.1 & 4.3.4) coupled with the deep learning TSP algorithm (3.3.3) to stratify patients for the Immune and Stromal subtypes. Given that Ki67 stained tissue was only available in the form of TMAs not full sections, the Ki67 %PI was determined for each patient from the TMA according to the methodology described in 5.3.1 and the resulting scores were used to stratify the patients for the Canonical and Latent subtypes post hoc within the SPSS database.

The two methodologies for conducting the Phenotypic Subtype assessment via image analysis were initially validated in the Glasgow Development cohort. Kaplan – Meier survival analysis of the Phenotypic Subtypes determined through semi – automated image analysis showed a statistically significant association with CSS, performing comparably to manual histopathological assessment (HR 1.528 95% CI 1.236-1.888 $P < 0.0001$ vs HR 1.545 95% CI 1.263-1.890 $P < 0.0001$, respectively) (Figure 6.9).

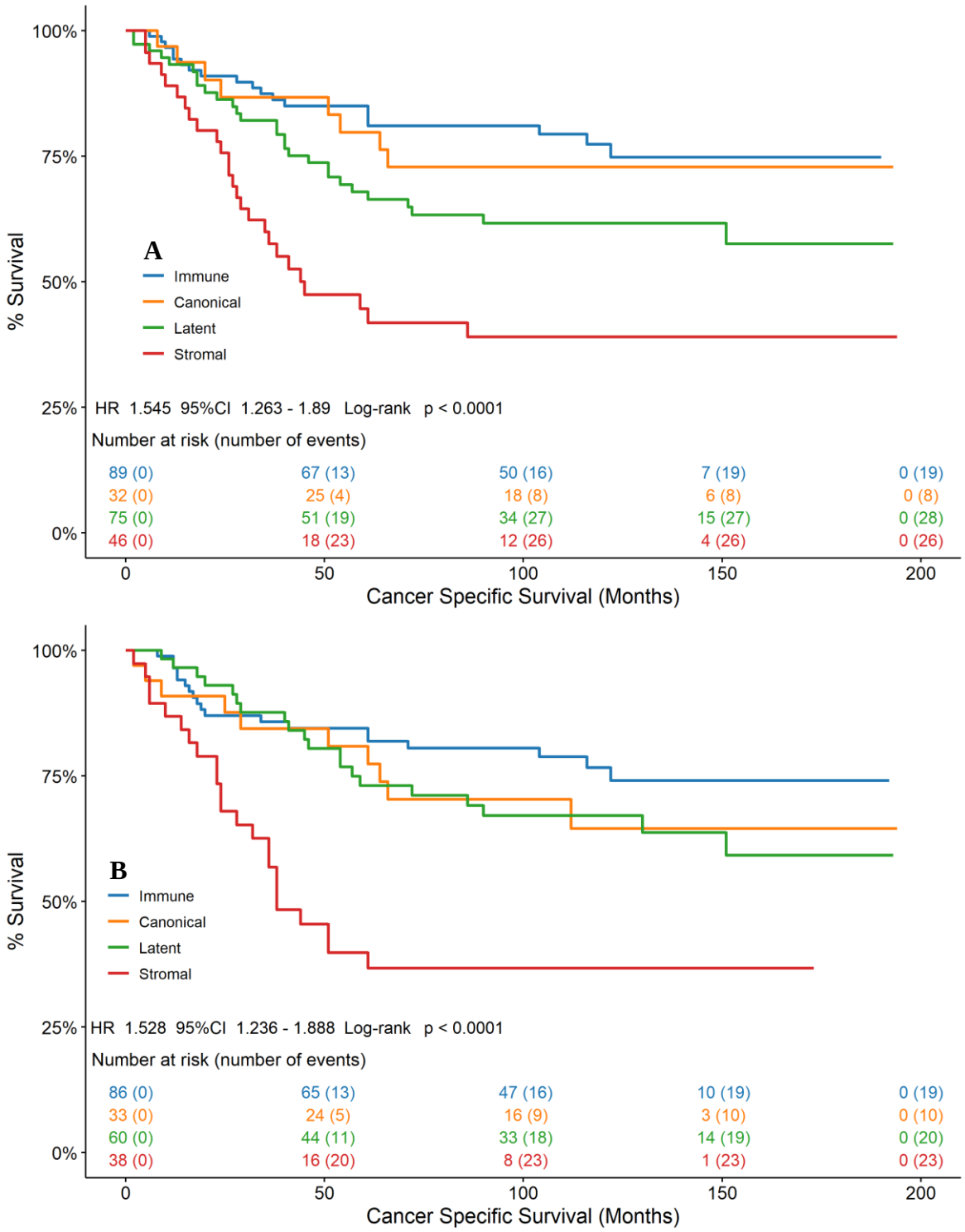


Figure 6.9 Relationship Between CSS and Phenotypic Subtype in Glasgow Development Cohort. *Association of CSS and Phenotypic Subtype determined via manual histopathological assessment (A) and semi – automated image analysis assessment (B) in the Glasgow Development Cohort.*

Kaplan – Meier survival analysis of the fully automated method for conducting Phenotypic Subtype assessment also showed a statistically significant association with CSS in the Glasgow Development cohort, though generating a notably lower hazard ratio and statistical significance than manual histopathological assessment (HR 1.345 95% CI 1.105-1.638 P = 0.00072 vs HR 1.545 95% CI 1.263-1.890 P < 0.0001, respectively) (Figure 6.10).

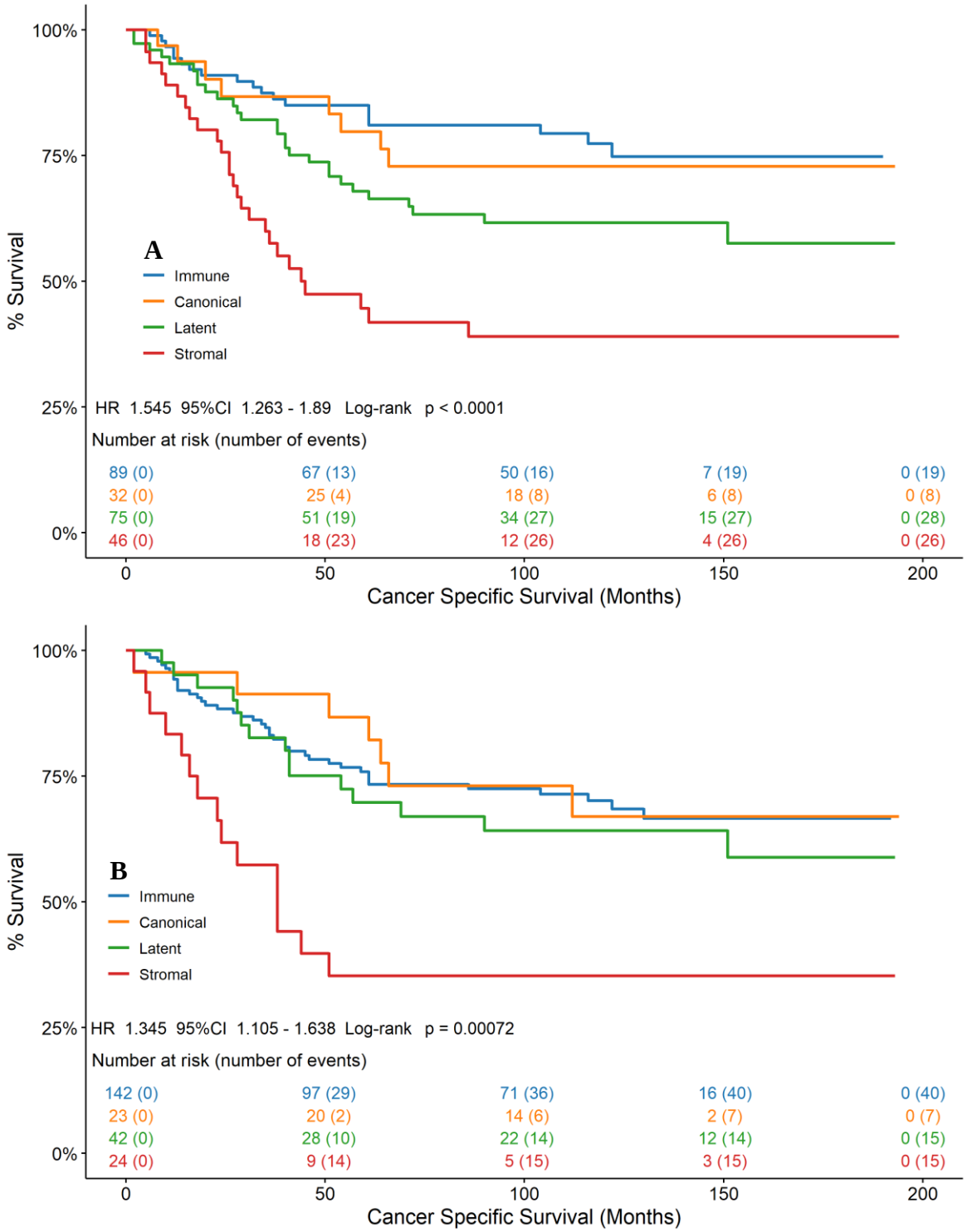


Figure 6.10 Relationship Between CSS and Phenotypic Subtype in Glasgow Development Cohort. *Association of CSS and Phenotypic Subtype determined via manual histopathological assessment (A) and fully automated image analysis assessment (B) in the Glasgow Development Cohort.*

6.3.5 Validation of Image Analysis Phenotypic Subtype Assessment in Glasgow Validation Cohort

Having demonstrated that the image analysis approaches were statistically significantly associated with CSS in the Glasgow Development cohort, the two approaches were then validated for prognosis in the Glasgow Validation cohort. The TSP and lymphocyte detections algorithms were applied to the H&E sections identified for image analysis in 4.3.2 and the Ki67 assessment was conducted on TMA. As per 6.3.4, the H&E TSP and immune infiltrate scores and TMA Ki67 scores were used to stratify patients into the Phenotypic Subtypes post hoc in the SPSS database.

Kaplan – Meier survival analysis of the Phenotypic Subtypes determined via semi – automated image analysis showed a statistically significant association with CSS; however, this association was notably less significant than manual histopathological assessment (HR 1.267 95% CI 1.097-1.463 P = 0.0026 vs HR 1.466 95% CI 1.247-1.724 P < 0.0001, respectively) (Figure 6.11).

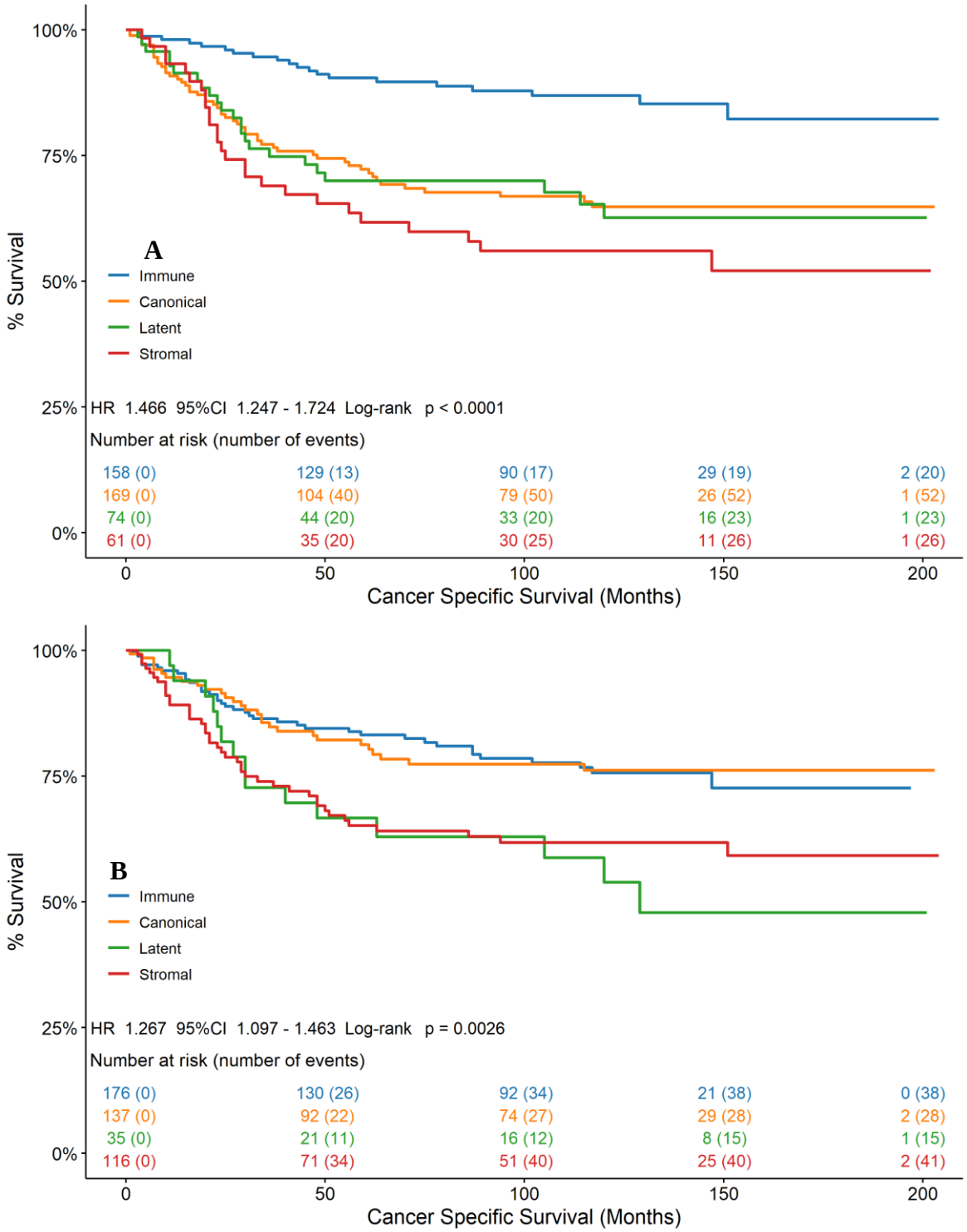


Figure 6.11 Relationship Between CSS and Phenotypic Subtype in Glasgow Validation Cohort. *Association of CSS and Phenotypic Subtype determined via manual histopathological assessment (A) and semi – automated image analysis assessment (B) in the Glasgow Validation Cohort.*

Kaplan – Meier survival analysis of the fully automated image analysis approach to the Phenotypic Subtypes also showed a statistically significant association with CSS in this cohort, again however, the significance of this association was notably lower than that of manual histopathological assessment (HR 1.236 95% CI 1.077-1.419 P = 0.0033 vs HR 1.466 95% CI 1.247-1.724 P < 0.0001, respectively) (Figure 6.12).

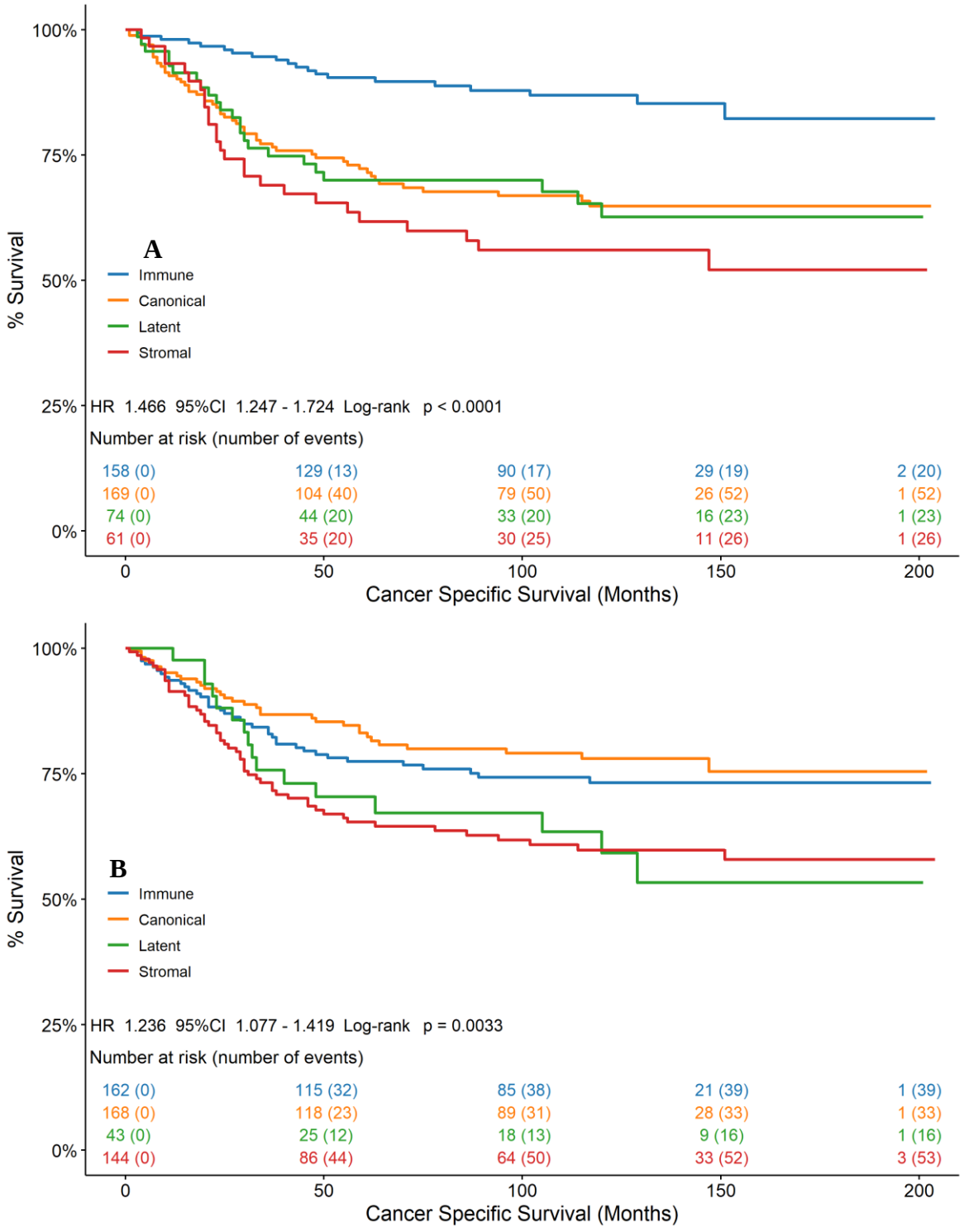


Figure 6.12 Relationship Between CSS and Phenotypic Subtype in Glasgow Validation Cohort. *Association of CSS and Phenotypic Subtype determined via manual histopathological assessment (A) and fully automated image analysis assessment (B) in the Glasgow Validation Cohort.*

6.3.6 Image Analysis of Phenotypic Subtypes and Association with Chemotherapeutic Response in TransSCOT Clinical Trial Cohort

Final validation of the image analysis approaches to the Phenotypic Subtype assessment was conducted in the TransSCOT clinical trial cohort. As detailed in 5.3.4, Ki67 assessment in this cohort was originally conducted on 1317 full sections as part of a study validating the Phenotypic subtypes and its components for prognostic significance, however, image analysis was performed on the TMA constructed from this cohort in the interest of consistency of assessment with previous cohorts and extracting as much clinically relevant data as possible, as the TMA was constructed from 2352 patients. Therefore, the manual histopathological assessment and image analysis could be considered too methodologically distinct to interpret information accurately and reliably from any measures of statistical comparison and / or agreement between the two, hence why no such analysis was conducted in Chapter 5. However, in the present study, it is of benefit for the validation of the image analysis approaches to the Phenotypic Subtypes to conduct a heuristic comparison between the two methods and the manual histopathological assessment. To perform this, the two image analysis algorithms were applied to the H&E WSI and Ki67 TMA core to stratify patients into their respective Phenotypic Subtype. Kaplan – Meier survival analysis was then performed on patients that had both an image analysis assigned and manually assigned Phenotypic Subtype for comparison, and on all available patients for the image analysis approaches.

Survival analysis of the patients with an image analysis and manually assigned Phenotypic Subtype showed that semi – automated approach demonstrated a statistically significant association with DFS, outperforming manual histopathological assessment (HR 1.261 95% CI 1.135-1.401 P = 0.00021 vs HR 1.197 95% CI 1.086-1.320 P = 0.0041, respectively) (Figure 6.13). The fully automated approach on these same patients showed an elevated hazard ratio and an association with DFS trending toward significance (HR 1.16 95% CI 1.041-1.293 P = 0.059) (Figure 6.14). Survival analysis of all patients available for image analysis demonstrated that both Subtyping methods have a statistically significant association with DFS, with the semi – automated approach showing a stronger association than the fully automated approach (HR 1.332 95% CI 1.219-1.454 P < 0.0001 vs HR 1.227 95% CI 1.113-1.328 P < 0.0001, respectively) (Figure 6.15).

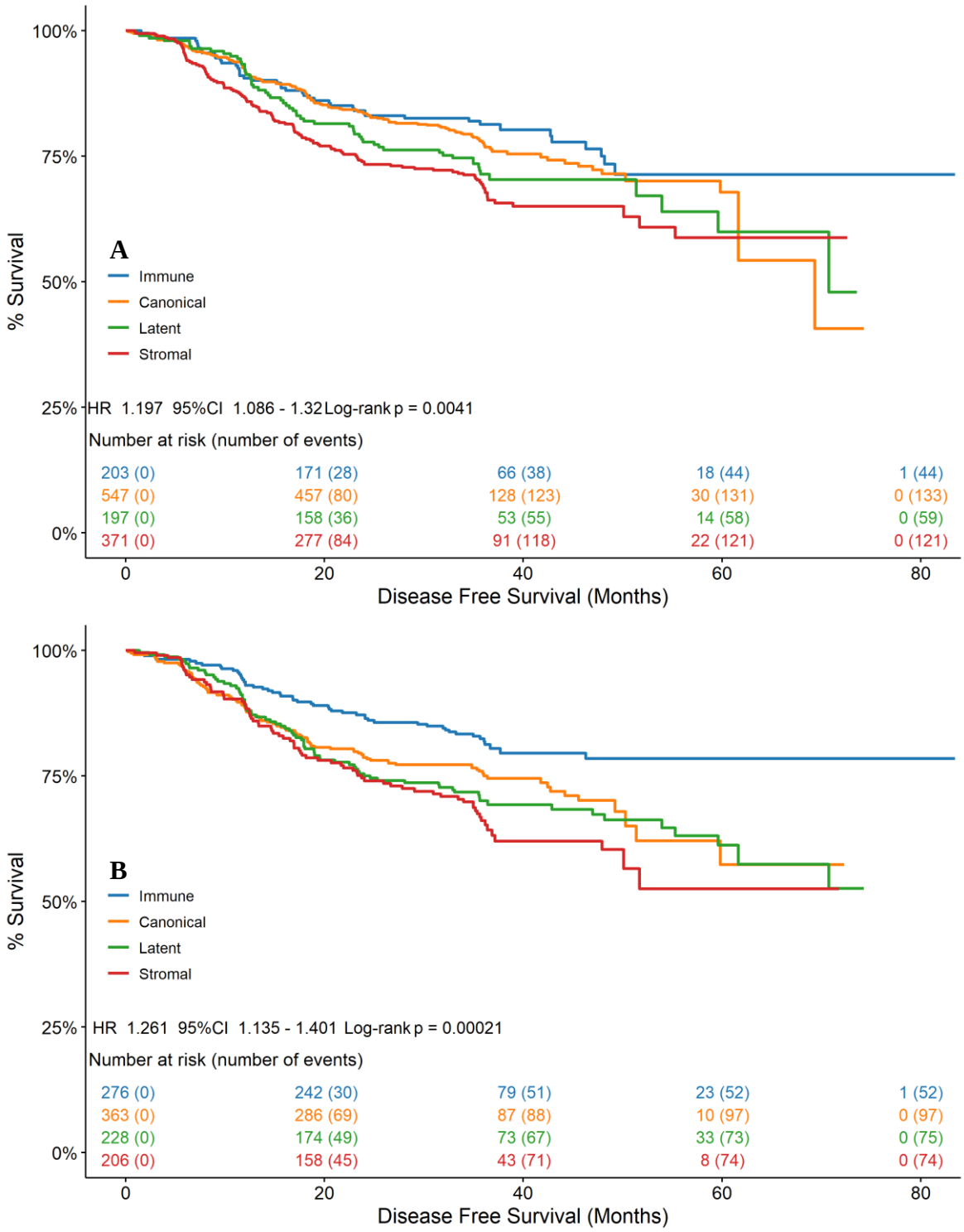


Figure 6.13 Relationship Between DFS and Phenotypic Subtype in TransSCOT Clinical Trial Cohort. *Association of DFS and Phenotypic Subtype determined via manual histopathological assessment (A) and semi – automated image analysis assessment (B) in the TransSCOT Clinical Trial Cohort.*

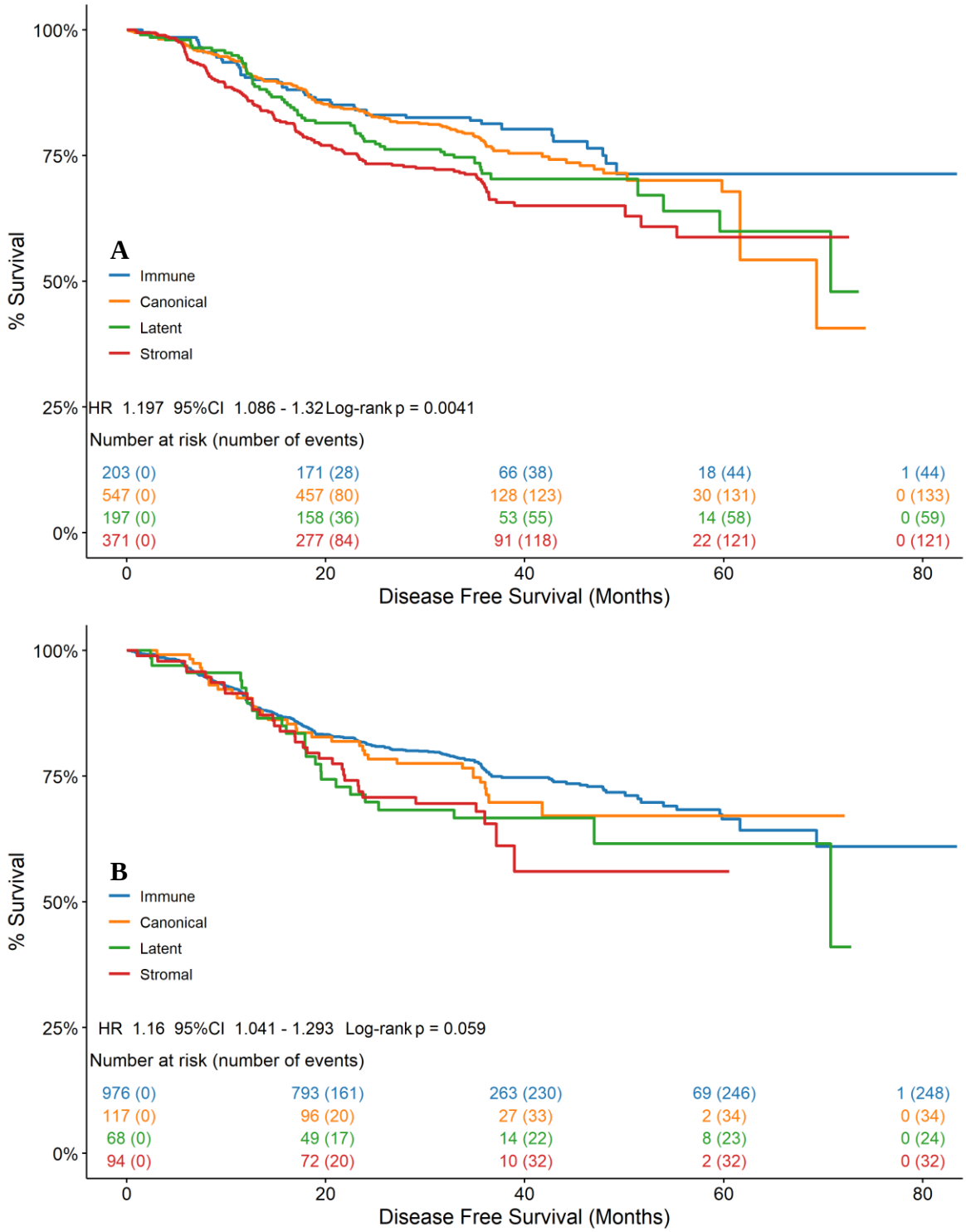


Figure 6.14 Relationship Between DFS and Phenotypic Subtype in TransSCOT Clinical Trial Cohort. *Association of DFS and Phenotypic Subtype determined via manual histopathological assessment (A) and fully automated image analysis assessment (B) in the TransSCOT Clinical Trial Cohort.*

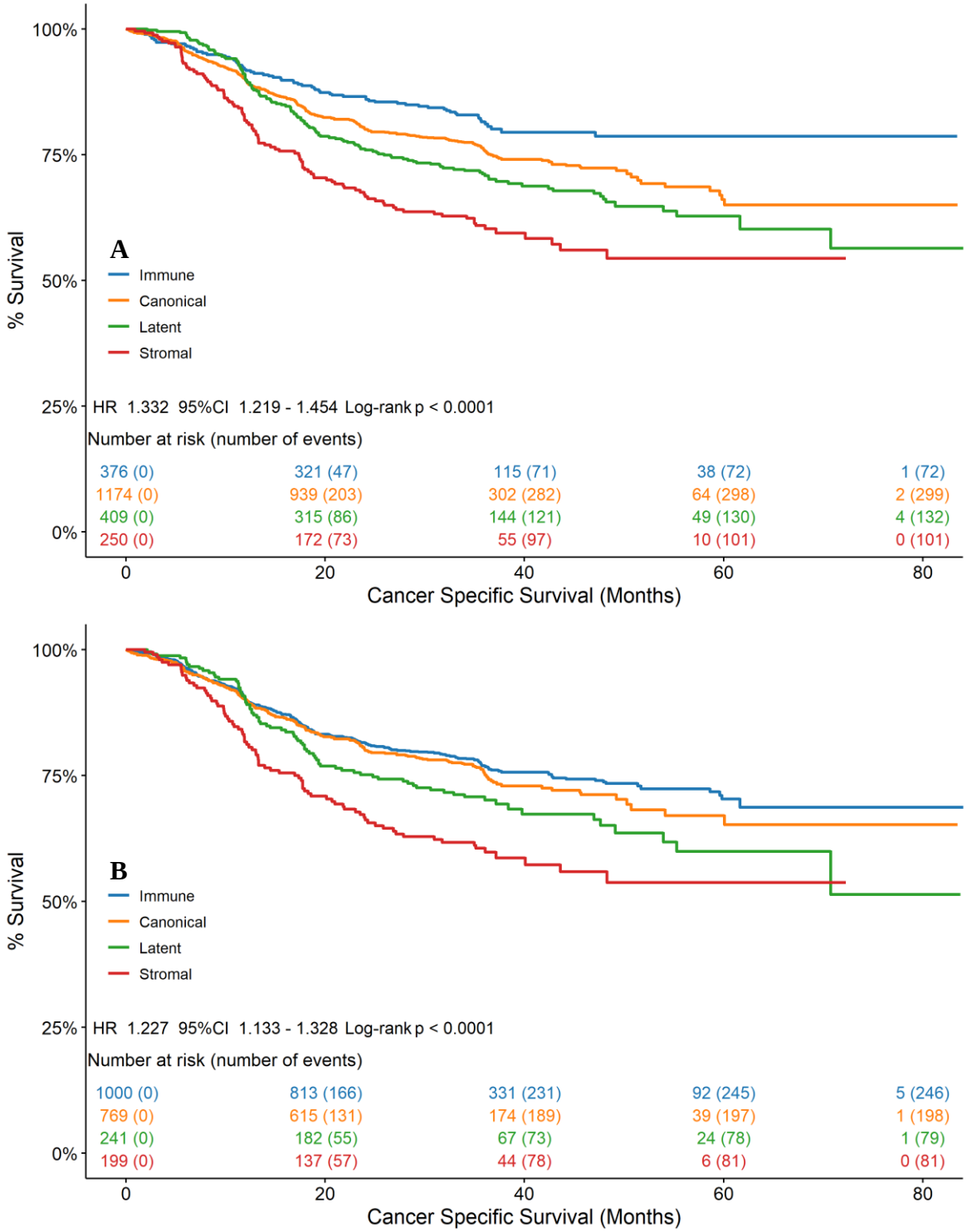


Figure 6.15 Relationship Between DFS and Phenotypic Subtype in TransSCOT Clinical Trial Cohort. *Association of DFS and Phenotypic Subtype determined via semi – automated (A) and fully automated image analysis assessment (B) in the TransSCOT Clinical Trial Cohort.*

As it was the most prognostic Subtype model, the semi-automated algorithm underwent univariate and multivariate Cox regression analysis to determine the relationship with DFS and prognostic independence from other clinicopathological features. On multivariate analysis, the semi-automated GMS algorithm was found to be significantly associated with DFS (HR 1.776 95%CI 1.360-2.321, $p < 0.001$), independent of T-Stage ($p < 0.001$) and N-Stage ($p < 0.001$), Table 6.4.

	Univariate				**Multivariate**		
	HR	95% CI	p-value		HR	95% CI	p-value
Sex							
Female	1.0						
Male	1.040	0.898-1.205	0.599				
T Stage				**T Stage**			
1	1.0			1	1.0		
2	1.013	0.498-2.060	0.973	2	1.205	0.490-2.963	0.684
3	1.85	0.988-3.467	0.055	3	2.088	0.930-4.691	0.074
4	3.217	1.715-6.037	**<0.001**	4	3.760	1.669-8.471	**<0.001**
N Stage				**N Stage**			
0	1.0			0	1.0		
1	1.436	1.155-1.787	**0.001**	1	1.914	1.508-2.430	**<0.001**
2	2.821	2.249-3.538	**<0.001**	2	3.253	2.553-4.146	**<0.001**
Site				**Site**			
Left	1.0			Left	1.0		
Right	0.687	0.557-0.848	**<0.001**	Right	0.817	0.647-1.032	0.089
Treatment							
FOLFOX	1.0						
CAPOX	1.084	0.924-1.272	0.319				
Phenotypic Subtype				**Phenotypic Subtype**			
Immune	1.0			Immune	1.0		
Canonical	1.414	1.116-1.791	**0.004**	Canonical	1.294	1.021-1.640	**0.033**
Latent	1.720	1.318-2.245	**<0.001**	Latent	1.535	1.174-2.006	**0.002**
Stromal	2.042	1.567-2.663	**<0.001**	Stromal	1.776	1360-2.321	**<0.001**

Table 6.4 Relationship between Phenotypic Subtype, Clinicopathological Features and DFS in TransSCOT Cohort. Hazard ratios and 95% confidence intervals determined via Cox proportional hazards regression and multivariate analysis conducted using backwards conditional method.

In the study validating the Phenotypic Subtypes via manual histopathological assessment in the TransSCOT clinical trial cohort, Roseweir and Colleagues identified that the Immune subtype had a statistically significant improvement in survival when administered Folinic Acid + 5 – Fluorouracil + Oxaliplatin (FOLFOX) chemotherapy over Capecitabine + Oxaliplatin (CAPOX) (HR 1.67 P = 0.019), an effect that was not present in the other subtypes (Roseweir et al., 2020). To evaluate whether this effect could be recapitulated through image analysis, patients assigned a Phenotypic Subtype through both image analysis methods were stratified by chemotherapy regimen and Kaplan – Meier survival analysis was conducted on all four subtypes from both methods.

The Immune subtype patients determined through semi – automated image analysis demonstrated a non – statistically significant association of chemotherapy regimen with DFS, however, the hazard ratio was comparable to that determined in the original study (HR 1.634 95% CI 0.909-2.935 P = 0.09) (Figure 6.16). Additionally, no other subtype determined via this image analysis method showed a trend towards a significant association with DFS (Figure 6.16).

In the Phenotypic Subtypes determined via fully automated image analysis, the Immune subtype did not demonstrate any notable association between DFS and chemotherapy regimen (HR 0.942 95% CI 0.750-1.183 P = 0.61, Figure 6.17), however, the Canonical subtype unexpectedly showed a statistically significant association between DFS and chemotherapy regimen (HR 1.526 95% CI 1.010-2.306 P = 0.043, Figure 6.16). Additionally, neither the Latent or Stromal subtypes showed any association between DFS and chemotherapy regimen (Figure 6.17).

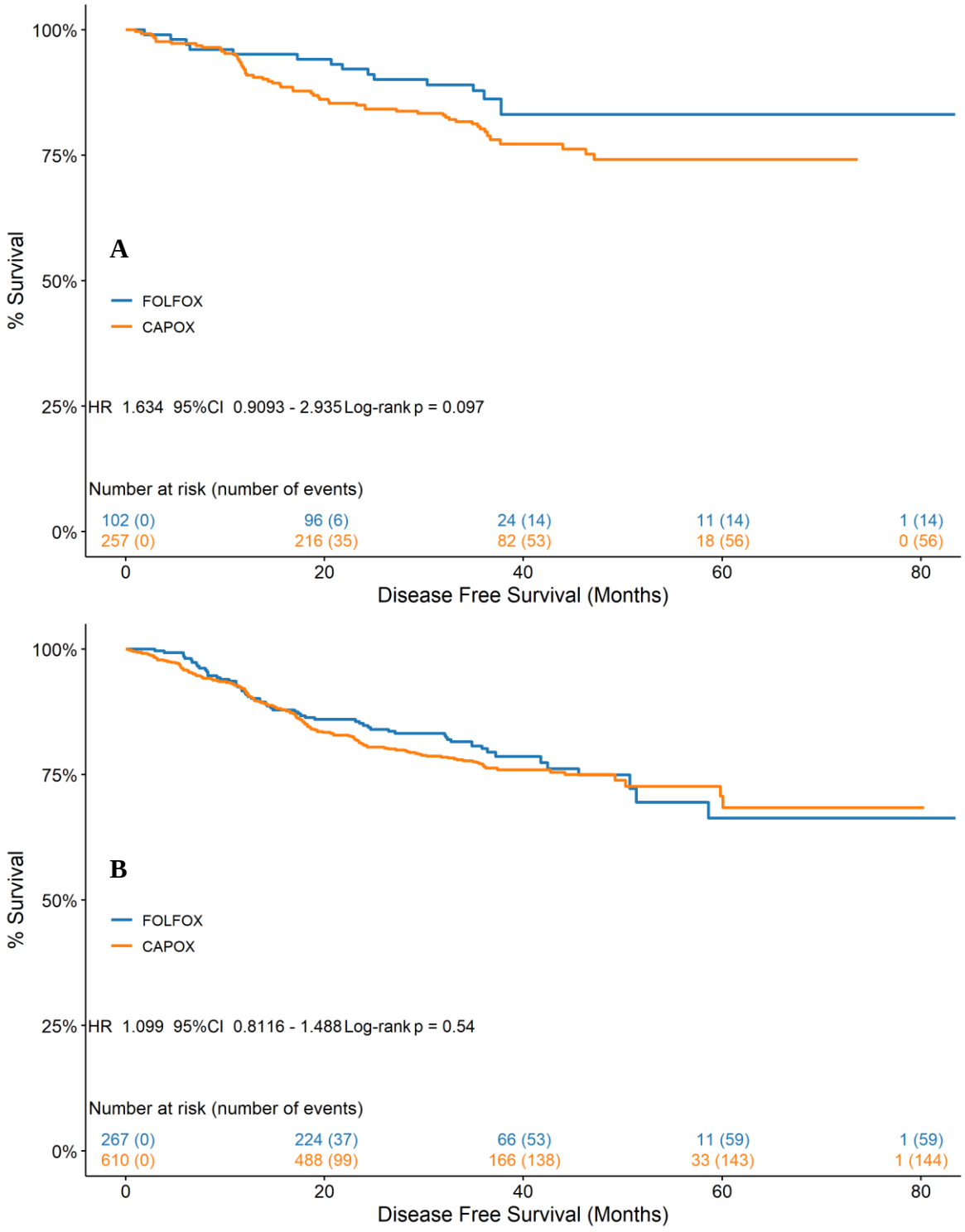


Figure 6.16 Continued on next page.

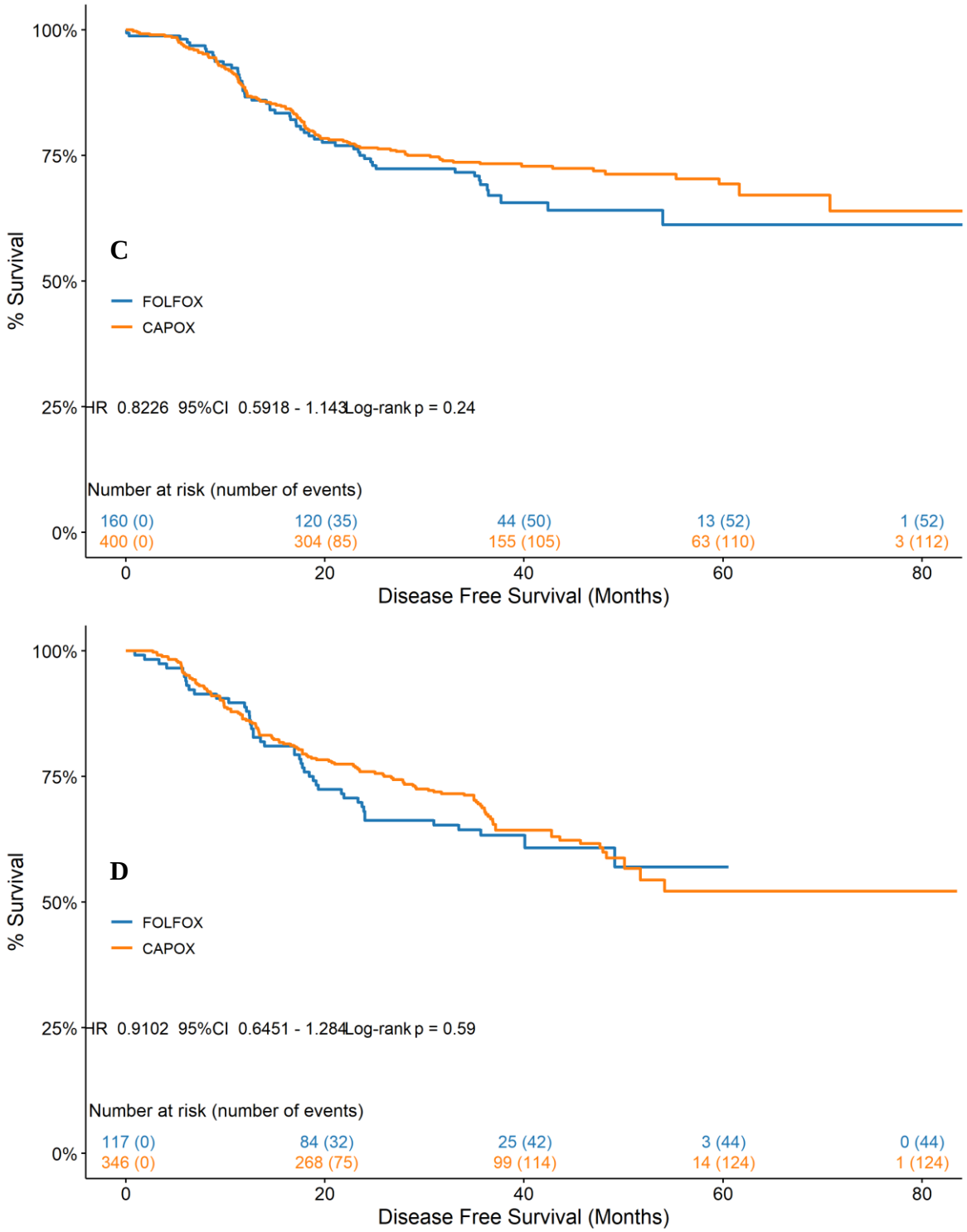


Figure 6.16 Relationship Between DFS and Chemotherapy Regimen in TransSCOT Clinical Trial Cohort. *Association of DFS and Chemotherapy Regimen in the Immune (A), Canonical (B), Latent (C), and Stromal (D) Phenotypic Subtypes determined via semi – automated image analysis assessment (B) in the TransSCOT Clinical Trial Cohort.*

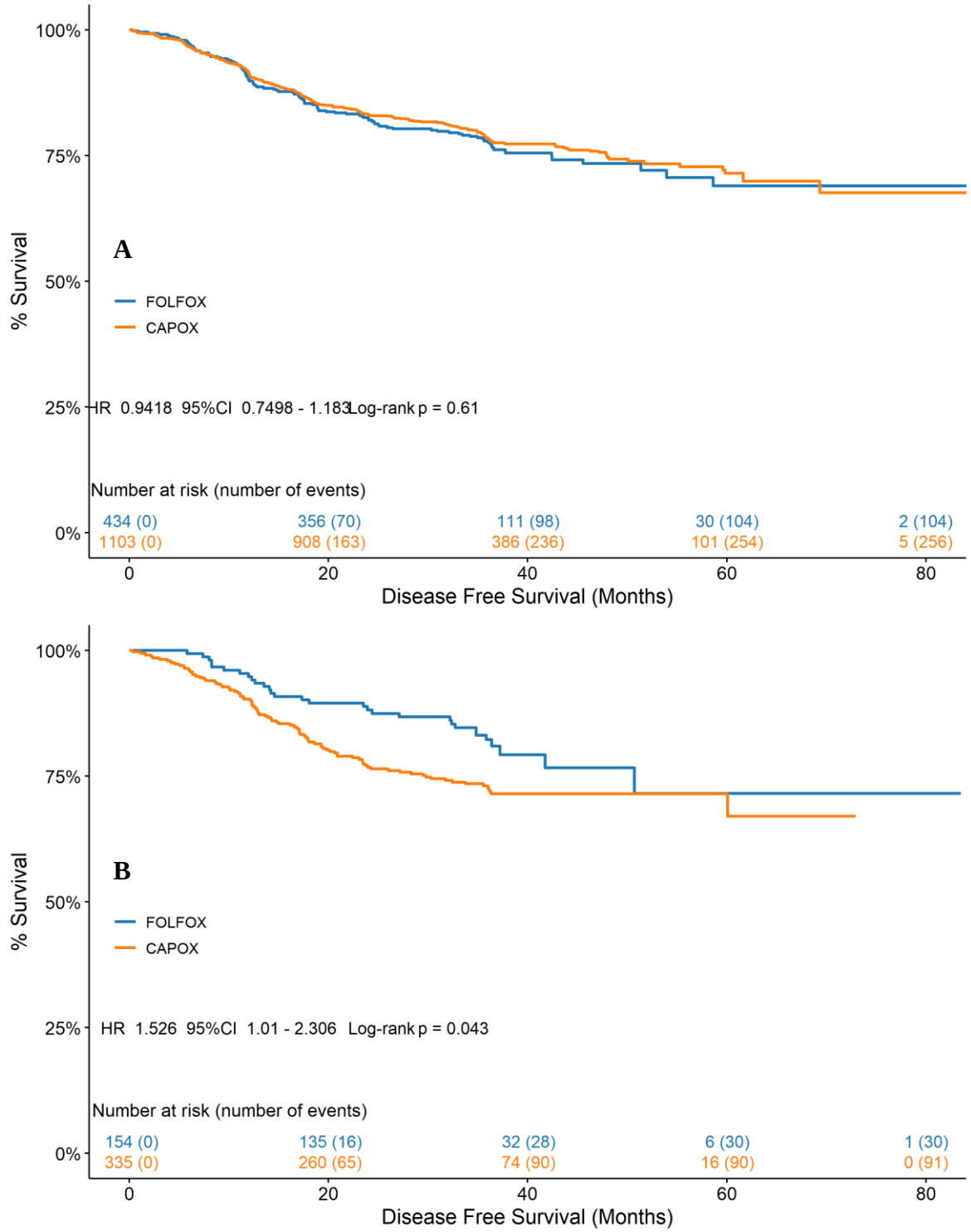


Figure 6.17 Continued on next page.

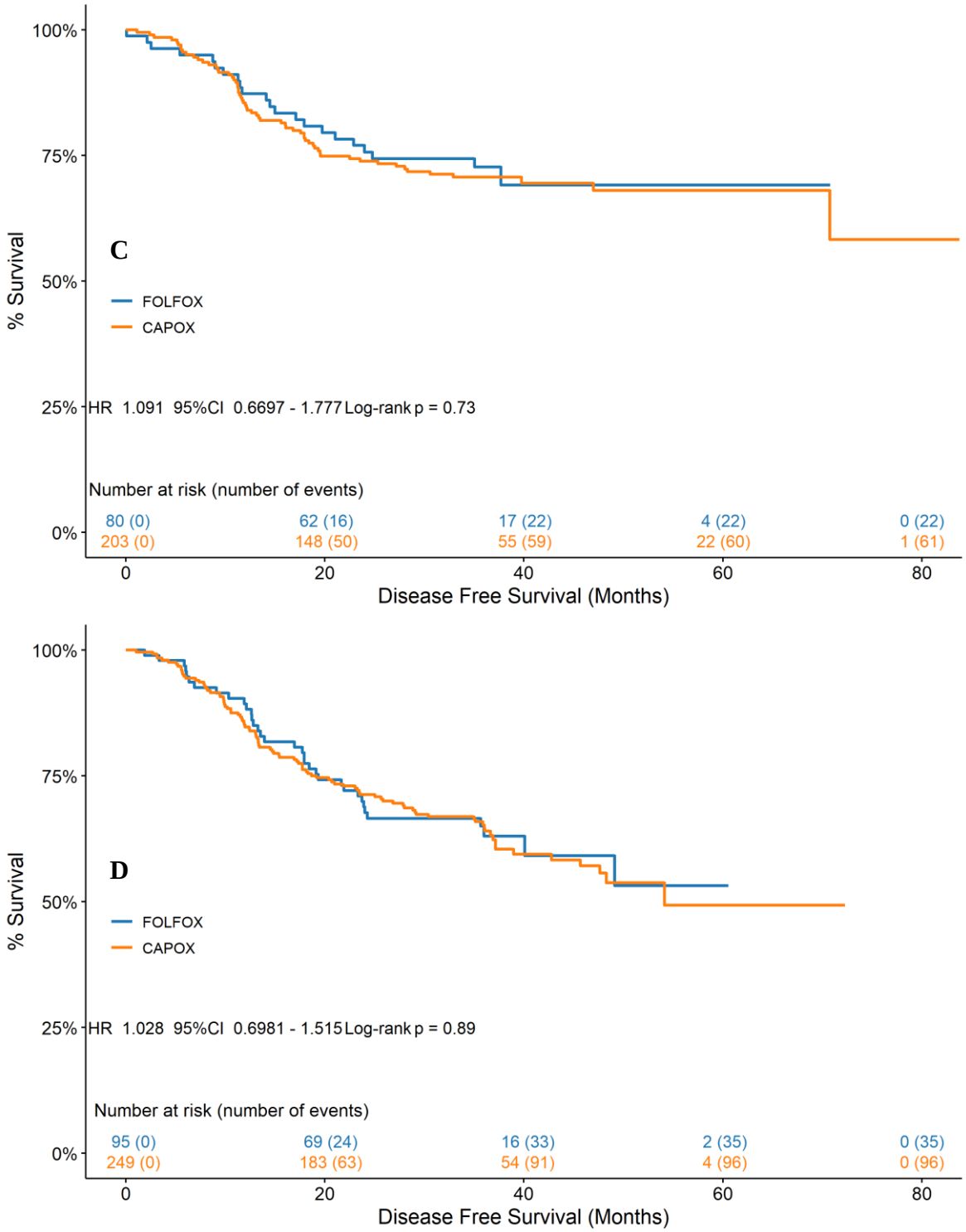


Figure 6.17 Relationship Between DFS and Chemotherapy Regimen in TransSCOT Clinical Trial Cohort. *Association of DFS and Chemotherapy Regimen in the Immune (A), Canonical (B), Latent (C), and Stromal (D) Phenotypic Subtypes determined via fully automated image analysis assessment (B) in the TransSCOT Clinical Trial Cohort.*

6.4 Discussion

The persistent issue of determining best clinical practices for Stage II patients in terms of administering adjuvant chemotherapy and follow up regimens presents a continually unmet clinical need in the TNM staging criteria. Multiple studies have investigated the possible implementation of novel classification systems determined through a variety of modalities to address this issue. Early work by Jass and Colleagues sought to use novel pathological features, such as the peritumoural inflammatory response and characterisation of the morphology of the invasive margin, to provide a more complete picture of the pathology of the CRC TME (Jass et al., 1987). While this system was never taken forward to clinical practice due to a lack of reproducibly standardised criteria, it set the precedent of combining novel prognostic features into a classification system to better assess the outcomes of specific CRC cases. Petersen and Colleagues expanded upon this work in 2002 utilising peritoneal involvement, venous invasion, marginal involvement, and tumour perforation in a combinatorial score that could identify high – risk Stage II patients (Petersen et al., 2002). However, despite incorporating features that are routinely assessed in clinical practice and addressing the issue with the TNM criteria, this classification system again saw no clinical implementation since its inception.

Advancements in genomic technologies saw a proliferation of studies using shared gene expression signatures of CRC to identify subtyping systems. From these studies, six of the most well characterised subtyping methodologies were amalgamated into a single combined subtyping system through clustering analysis by Guinney et al., 2015. This combined subtyping system, known as the Consensus Molecular Subtypes (CMS), identified four independent subtypes based on shared genetic and phenotypic characteristics across the six precursor systems. While this methodology showed that each subtype conferred an individual prognosis and could potentially direct targeted chemotherapy regimens, the expensive implementation of the genomic techniques used to generate the subtypes coupled with concerns over the masking effect of stromal – derived gene signatures in CMS4 meant that this too was not adopted clinically.

Through a series of studies examining features of the CRC TME assessable on diagnostic H&E sections, Park and Colleagues identified the KM system of peritumoural inflammatory response grading and quantification of the tumour – associated stroma through the TSP as possible candidates for inclusion in a combinatorial assessment (Park et al., 2014b, 2015a). The Glasgow Microenvironment Score (GMS) is an independently prognostic classification system grouping patients into three subtypes that has since been validated in

multiple independent cohorts (Alexander et al., 2021a). However, the subjective assessment criteria of its components has hindered its progress to clinical application. Following this work, Roseweir and Colleagues conducted a study aiming to translate the phenotypic characteristics of the CMS to routine histopathology, highlighting that the prominent inflammatory response of CMS1 correlated with GMS0 and as such could be identified through the KM criteria, and similarly the stromal infiltrate of CMS4 correlated with GMS2 and could therefore be identified through the TSP criteria (Roseweir et al., 2017). Combining these two H&E – based assessments with Ki67 IHC to identify the CMS2 proliferative phenotype through histology, resulted in the Phenotypic Subtypes, a four group independently prognostic classification system that has been validated in independent cohorts and importantly, appears to predict response to chemotherapy in immunologically hot tumours (Roseweir et al., 2018b, 2020).

The Phenotypic Subtypes possess the same issues of subjectivity in assessment criteria as the GMS, and as such, have also yet to see use in clinical practice. Therefore, applying image analysis to these classification systems could ensure the reproducibility needed for clinical adoption. To perform these assessments digitally, the image analysis algorithms for TSP quantification and lymphocyte infiltrate described in Chapter 3 and Chapter 4, respectively, were performed on a single H&E section to determine the GMS, and coupled with automated Ki67 assessment, described in Chapter 5, *post hoc* to determine the Phenotypic Subtypes.

As described in Chapter 4, two approaches were used to quantify lymphocyte infiltrate, peritumourally through semi – automated image analysis and intratumourally through automated image analysis, therefore, both approaches were incorporated into the GMS and Phenotypic Subtype assessments to assess their interaction with the TSP and resulting prognostic significance. The algorithms were first validated in the Glasgow Development cohort where both methods showed significant associations with CSS and stratified patients for prognosis following the same pattern as manual histopathological assessment, with the high immune group GMS0 possessing the best prognosis, the stromally dense GMS2 the worst prognosis, the low for both group GMS1 intermediate prognosis (Figure 6.2 & 6.4). The semi – automated approach demonstrated the best association with CSS in terms of hazard ratio, outperforming manual histopathological assessment, and while there was a notable drop in association with CSS with the automated approach, it still retained statistical significance and indeed, this was to be expected as intratumoural assessment of lymphocytes on H&E is prognostically inferior to peritumoural assessment

(Alexander et al., 2020). This prognostic association was confirmed in the Glasgow Validation cohort where similarly, both approaches showed a significant association with CSS however, neither approach outperformed manual histopathological assessment (Figure 6.5 & 6.6) and a marked drop in GMS0 prognosis was associated with automated assessment though again still retaining significance (Figure 6.6). The discrepancy in prognostic significance between manual histopathological assessment and image analysis in this cohort is likely due to the issues in section quality and discordance between the sections used to conduct each assessment described in 3.4 and 4.4, however, it is promising that the algorithms still prognostically perform in a multi – group system despite these issues. This pattern was similarly seen in the TransSCOT clinical trial cohort, where the semi – automated approach performed almost identically to manual histopathological assessment for DFS prognosis (Figure 6.7), and the automated approach still retained a significant association with DFS despite the poorer GMS0 prognosis (Figure 6.8). The WSI from this cohort are the most recently produced and of the highest quality of all the cohorts investigated, thus lending credence to the hypothesis that intratumoural lymphocyte assessment on H&E is prognostically inferior to peritumoural assessment, and that TIL assessment should be conducted through IHC to subset the lymphocytes and delineate their individual prognostic effect.

Image analysis – based assessment of the Phenotypic Subtypes in the Glasgow Development cohort further strengthened the prognostic capability of digital pathology in multi – group grading systems. The semi – automated approach showed a near identical association with CSS to that of manual histopathological assessment and the automated approach still retained significance despite the predicted drop in prognosis for the Immune subtype. The discordance in the prognostic significance for the Subtypes through automated assessment is potentially due to the skew of patients to the Immune subtype produced through TIL assessment instead of peritumoural assessment. In the Glasgow Validation cohort, the semi – automated assessment shows some Immune subtype preference compared to manual histopathological assessment, but still broadly retains the same prognostic pattern. Interestingly, the patient distributions achieved through automated assessment in this cohort are similar to manual assessment, however, the Immune subtype possesses a poorer prognosis than the Canonical subtype (Figure 6.12), suggesting that in the absence of peritumoural inflammatory assessment, the proliferative activity of the tumour becomes the most prognostically significant factor.

In the TransSCOT Clinical Trial Cohort, the semi – automated approach to the Phenotypic Subtypes demonstrated a markedly better association with DFS than manual assessment. However, when utilising the fully automated assessment, the Immune subtype preference was again present, heavily impacting the association with DFS resulting in a non – significant prognosis. It is of note that the Canonical subtype appears to have a marginally better prognosis than the Immune subtype as was seen in the Glasgow Development cohort. However, image analysis subtyping in all available patients appears to restore the prognostic pattern in the automated assessment and the Immune subtype retains the best prognosis despite patient numbers again being skewed towards this subtype. The semi – automated approach again demonstrated the best association with DFS and produced a more marked prognostic difference between the subtypes.

Roseweir and Colleagues further demonstrated that Immune subtype patients had a significantly improved outcome when administered FOLFOX over CAPOX chemotherapy in the TransSCOT cohort (Roseweir et al., 2020). While not significant, the Immune subtype determined through semi – automated assessment did show a trend towards a significant difference in prognosis between the two chemotherapy regimens (Figure 6.16 A), and indeed this was the only subtype to show any association between survival and chemotherapy regimen (Figure 6.16 B – D). Through automated analysis, the Canonical subtype demonstrated a significant improvement in survival for patients administered FOLFOX (Figure 6.17 B), where no association was seen in any other subtype (Figure 6.17 A, C – D). This could again suggest that tumour proliferation is the most important prognostic factor in the absence of peritumoural lymphocyte assessment and that there is a fundamental biological difference in lymphocyte response at the tumour's invasive margin compared to an intratumoural response. It would be of interest to investigate whether lymphocyte subset analysis is able to define the response to chemotherapy from an intratumoural perspective and if so, whether it is just the means of assessment that defines this response or if a specific lymphocyte subtype is modulating the efficacy of the chemotherapy *in situ*.

While this study shows the prognostic significance of image analysis, it falls short on some requirements of a clinically viable system. Notably the need to annotate the invasive margin to appropriately determine GMS0 / Immune subtype in order to recreate the nuances seen with this group of patients through manual histopathological assessment. This is not a particularly laborious task for an experienced pathologist; however, it still requires additional time and is a somewhat subjective task, which could be seen as sufficient to make it impractical for clinical use. As previously discussed, this could be addressed through deep

learning using a large training data set of annotated invasive margins to extract high level features that distinguish the most clinically relevant point of tumour invasion. Additionally, the Ki67 portion of this assessment was performed using TMA and was not validated for use in WSI, however, the means by which to do this through automation exists within the Visiopharm Tissuealign module, whereby the tumour class masks could be overlayed on a serially cut, Ki67 IHC section to distinguish the tumour, and then apply the previously validated Ki67 positive cell detection algorithm within these areas. This would further the clinical utility of the image analysis approaches as it would only necessitate an additional section to be cut, stained, and scanned, given that the immune infiltrate can be adequately assessed from the diagnostic H&E WSI via the KM criteria. However, if the assessment was to be conducted on a different histological modality, such as biopsies, then it would be more clinically appropriate to cut a second section for the purpose of CD3 staining. This is due to the fact that assessment of the immune infiltrate conducted on biopsy H&Es does not hold sufficient prognostic significance when compared to lymphocyte subtyping (Park et al., 2019).

In summary, this chapter clearly demonstrates that image analysis algorithms operating on multiple tissue modalities can be successfully applied in combination to accurately determine classification systems with multiple, individually prognostic groups, in independent clinical cohorts. Additionally, biological phenomena determined through histopathology can be recreated through image analysis with careful and methodical application to appropriate WSI. Future work focussing on reducing the input required from a pathologist prior to analysis while retaining the clinically relevant information produced by the algorithms would ensure the clinically translatability of image analysis approaches to these prognostic systems, providing important clinical information regarding treatment regimens, and ultimately improving patient outcomes.

7. Discussion

7.1 Discussion

CRC is one of the most common malignancies in the world, accounting for an estimated 10% of all new cancer cases and mortalities. Imperative to the continued improvement of care of CRC patients is adequate determination of prognosis and appropriate therapeutic regimens. Despite continuing research into the identification of novel prognostic features and more effective treatments, the TNM staging criteria which has served as the gold standard for prognosis and directing therapeutic decisions for nearly four decades, has been slow to incorporate new biomarkers that could address shortcomings in its clinical utility. The most well documented issue with the TNM criteria is in staging node negative disease and determining appropriate therapy strategies and follow up for these patients. Around 30% of Stage II patients experience notably poorer survival outcomes than the TNM stage identifies, and indeed low T – Stage, Stage III / node positive patients experience better survival outcomes than these patients, with a ~5% difference in 5 – year survival rates (Petersen et al., 2002; Sung et al., 2021). Despite the introduction of sub – staging to further stratify these patients based on clinical presentation and improve the prognostic accuracy, the identification of high – risk Stage II patients and guidance on administering appropriate therapy regimens persists as an unmet clinical need.

Studies identifying novel features of CRC holding prognostic significance that could supplement the TNM staging criteria date back nearly a century (Maccarty, 1931a), and a vast number of these features have been repeatedly validated in independent patient cohorts. One of the most well characterised factors known to modulate survival outcomes and response to chemotherapeutics in CRC, and the majority of other tumour types, is the host inflammatory response to the tumour. Indeed, it was this study by Maccarty in 1931 that first identified a prominent immune response to the tumour as a prognostically favourable factor as well as advocating for the grading of cancers based on the study of various criteria. The immunological response to CRC was further characterised in subsequent histopathological studies demonstrating the effect on prognosis and, in 1986, J.R. Jass introduced the first set of criteria for grading this response (Jass, 1986; C. S. D. Roxburgh & McMillan, 2012). Jass' work was further expanded upon by Klintrup and Colleagues in 2005, establishing a more comprehensive, four – part grading criteria for the perimtumoural lymphocyte infiltrate on H&E sections, which remains the most prognostic histological immune grading criteria and has been subsequently validated in multiple independent cohorts (Klintrup et al., 2005; Park, McMillan, et al., 2016; C. S. D. Roxburgh & McMillan, 2012; C. S. D. Roxburgh et al., 2009a, 2009b). Despite this, issues pertaining to interobserver variability have prohibited the clinical adoption of the KM grading system.

The relationship between the tumour and its microenvironment is one of the most important interactions governing survival outcomes. The tumour associated stroma has long been known to be the supporting tissue of the tumour, facilitating and modulating tumour growth and progression through a variety of heterogeneous processes, although isolating and interpreting its prognostic effect is a relatively recent development. Ueno and Colleagues initially identified a link between the stroma and prognosis through qualitative assessment of the type and maturity present in rectal tumours, showing that a more immature, less differentiated stroma conferred a significantly worse prognosis (Ueno et al., 2004; Ueno, Jones, et al., 2002b). Subsequent studies sought to investigate the amount of stroma present across the tumour area and establish a prognostic link, identifying the relative ratio of tumour to stroma, known as TSP / TSR, as the most effective way of quantifying the tumour associated stroma and stratifying patients for prognosis (Mesker et al., 2007, 2009). The TSP has since been prognostically validated in multiple CRC studies (Hynes et al., 2017c; Park, McMillan, et al., 2016; van Pelt et al., 2018b; West et al., 2010), multiple other cancer types (Kramer et al., 2019; Leppänen et al., 2019), and has been utilised to identify high – risk Stage II CRC patients that could benefit from adjuvant chemotherapy (Huijbers et al., 2013). Despite the breadth of supporting data, TSP assessment has also seen no implementation in clinical practice, however, an international reproducibility study now looking to standardise the assessment criteria for routine pathology (M. Smit et al., 2019).

IHC is routinely utilised in diagnostic pathology to assess the status of biomarkers that provide additional, molecular information regarding the nature of the tumour, for example MMR proficiency in CRC. The proliferative activity of the tumour is an important prognostic factor, and a hallmark of cancer, and histological measures of this activity have been clinically implemented in the diagnosis of other cancer types. In breast pathology, the biomarker Ki67 is used to distinguish between Luminal A and Luminal B molecular subtypes and guide chemotherapy regimens (Goldhirsch et al., 2013). In CRC, high tumoural Ki67 expression has been shown in multiple studies to be a favourable prognostic factor in primary disease (Fluge et al., 2009; Melling, Kowitz, Simon, et al., 2016; Reimers et al., 2014) and an unfavourable prognostic factor in metastatic disease (Nash et al., 2010a). Despite the availability of diagnostic Ki67 antibodies and optimization of the IHC protocol for assessing its expression, Ki67 has not been utilised in diagnostic pathology in CRC to date.

In Chapter 3, image analysis approaches to the TSP assessment were investigated for prognostic significance and translatability between patient cohorts. Although deep learning is becoming standard practice for tissue segmentation and classification in digital pathology, there is still a sizeable cost associated with its use, in the form of high – performance GPUs and associated IT infrastructure, that would prohibit its widespread use in remote areas and developing nations. While cloud computing may provide a potential solution to this, the security risk associated with the transfer of patient image data would preclude its use from settings such as clinical trials. Thus, in this study, an initial comparison between computationally less intensive, traditional machine learning algorithms and deep learning algorithms was conducted to identify the most appropriate means of conducting the TSP assessment through digital pathology.

The decision forest TSP algorithm was trained using hand crafted features, on a manually annotated subset (~5%) of the Norway patient cohort WSI to semantically segment the tissue and determine the TSP for each patient. In the full cohort of 299 patients, the algorithm was able to determine the TSP and significantly stratify all patients for CSS at a cut off value of 40%, performing comparably to manual assessment. Although manual histopathological assessment is stratified at 50% TSP, an optimized value was generated for the image analysis approach as absolute quantitation of the TSP had not previously been performed in any of the available patient cohorts. All previous assessment and associated TSP values had been determined via consensus between observers according to the criteria described by Park et al., 2014, and while this criterion directs the observer to estimate the TSP to the nearest 5%, it is still a subjective measure and previous studies have shown discrepancy between histopathological TSP assessment and image analysis (Zhao et al., 2020). It is also likely that analysing 50% of the image at random results in stochastic skewing of the TSP scores through not incorporating the whole tumour area in the assessment, whereas in manual histopathological assessment, an optimally representative area is chosen by the observer and thus this effect is not present.

Translating the decision forest algorithm to novel patient cohorts, the Glasgow Development and Validation cohorts, drastically reduced the prognostic ability of the algorithm and it did not significantly stratify patients for survival. This was somewhat expected as machine learning algorithms do not usually adapt well to novel data. This could be improved by including sections from a variety of cohorts in the training data, as it is likely that the algorithm learned feature relationships specific to the Norway cohort WSI due to the reliance on colour of the features used for classification. Certainly, it is possible that more

readily translatable TSP algorithm could be built using machine learning techniques coupled with a training data set that is broader in scope.

Utilising deep learning for this task resulted in a more prognostic and translatable algorithm. Data from the Norway and Glasgow Development cohorts show that deep learning is better suited to the task of semantic segmentation across multiple cohorts, as the image analysis outperformed manual histopathological assessment, which was performed by different observers, in both cohorts. The deep learning algorithm still translated significantly well to the Glasgow Validation Cohort however it did not perform as well as manual histopathological assessment. As previously stated, this is due to the quality of the WSI, which had been physically marked up by a pathologist thus the ink markings caused focussing issues during scanning. Additionally, there was notable variation in histological processing of the sections in this cohort, posing problems of variation in stain intensity and tissue artefacts, despite which the deep learning TSP approach was still able to significantly stratify patients for prognosis. Image analysis generally requires high quality WSI to perform optimally, and with modern histology labs are progressing towards digital pathology as the primary means of pathological reporting the quality of section processing and scanning is of the utmost importance. The routine clinical production of high – quality slides will inevitably act to bolster the application of image analysis to diagnostic use, as section quality will not need to be considered ensuring that the data produced by algorithms such as this will become increasingly more reliable. Evidence of this can be seen in data from the TransSCOT cohort, which produced the most recent and highest quality WSI. In this 2875 patient study, the deep learning algorithm again outperformed manual histopathological assessment for prognostic stratification.

One notable issue of this chapter is the variability of interobserver agreement between the image analysis and pathologist assessment, which raises the question of whether these metrics are appropriate for this type of study and indeed, to what extent they should be used to validate image analysis performance. In previous work where the TSP was conducted only by human observers, measures of statistical agreement were imperative to ensure consistency of assessment and validate that observers were not inadvertently taking other prognostic factors into account when assessing sections. If the ultimate aim of image analysis is to address the recurrent issues in interobserver variability, then perhaps another means of validating algorithm performance is required. It could be argued that the present study goes some way to providing a possible alternative to statistical agreement. An omnipresent issue with the use of artificial intelligence is that algorithms are often seen as black boxes, whereby

an input is given and an output is received without indication of the processes and features of importance used to determine the result. In this setting, particularly when metrics such as 5 – year survival outcomes are the only output, statistical measures of agreement are useful, if not imperative for ensuring algorithm reliability. However, with the approach and systems used in this study, the segmentation output of the algorithm can be mapped directly back to the section, allowing almost real time visualisation of the algorithms performance. It is conceivable therefore that in a clinical setting, a pathologist interested to determine the TSP of the current patient could apply the algorithm to the section, assess its performance, and ultimately decide whether to utilise the TSP as a prognostic or theragnostic factor in the pathological report.

Chapter 4 examined the translatability of the KM grading system of peritumoural inflammatory response to image analysis, in addition to automated tumour infiltrating lymphocyte assessment utilising the output from the TSP algorithm. Data from the Glasgow Development cohort identified a threshold – based traditional image analysis algorithm that was able to preferentially detect lymphocytes on clinical H&E sections. When this algorithm was applied within manually annotated invasive margins, the resultant lymphocyte densities significantly stratified patients for CSS prognosis, performing comparably to the manual KM assessment. Despite being a threshold – based algorithm derived from colour deconvolution features developed in the Glasgow Development cohort, this approach translated well to novel cohorts, significantly stratifying patients in both the Glasgow Validation and TransSCOT cohorts. The algorithm performed comparably to manual KM assessment in the TransSCOT cohort but was again limited by the section selection and WSI quality issues in the Glasgow Validation cohort. The prognostic significance of the KM assessment is likely linked to the morphological characteristics of the inflammatory response at the invasive margin, as these features appear to be ubiquitously conserved throughout CRC cohorts. The distinction in morphological characteristics is lost through simple quantitation of the lymphocyte density at the invasive margin. It is conceivable however, that a study investigating this phenomenon could elucidate a quantitative means of distinguishing a patchy vs band – like vs florid cup response, by examining the spatial relationships between the individual lymphocytes and between the lymphocytes and tumour cells, through techniques such as Euclidean distance matrix analysis or unsupervised clustering to distinguish patches of lymphocytes from contiguous populations. Additionally, more weakly supervised deep learning methods could be applied to translating the KM criteria more directly to image analysis. Using annotated examples of the individual KM grade patterns as training data for a neural network could elucidate features that distinguish

the morphological signatures of the lymphocyte responses themselves and predict the KM grade without the need for cellular quantification.

Stromal TIL quantification through automated image analysis also demonstrated prognostic significance in the Glasgow Development and TransSCOT cohorts, though failed to do so in the Glasgow Validation cohort. Quantification of TIL populations through image analysis is usually performed on sections stained for lymphocyte subset markers via IHC, as specific lymphocyte populations are most prognostic in this setting. This is evident through the comparison of H&E based TIL assessment with CD3 stained sections from the same cohorts, which show a greater association with survival. Deep learning has demonstrated proficiency in identifying cell types on H&E sections, such as tumour, fibroblast, immune, and parenchymal cells, following training on exhaustively annotated data sets. It is therefore possible that a large number of CRC sections stained for lymphocyte subsets could be used as ground truth for antibody directed learning, and high dimensional information extracted by a neural network be used to distinguish lymphocyte sub – populations based only on morphological features. This approach has seen some implementation in breast cancer, where pan – leukocyte CD45 IHC guided learning distinguished lymphocyte rich and poor areas when applied to H&E sections (Turkki et al., 2016).

In Chapter 5, the prognostic effect of Ki67 expression was analysed through image analysis. As described in Chapter 3, deep learning is not readily implemented everywhere, thus an initial comparison between a traditional cell classification algorithm and a commercially available deep learning nuclei detection algorithm adapted to detect Ki67 positive cells was conducted for prognostic significance. While both image analysis algorithms were highly correlated with manual $Ki67^{+}$ cell counting through statistical measures of agreement, the deep learning algorithm performed better for prognostic stratification for CSS in both the Glasgow Development and Validation cohorts, however only the Glasgow Validation cohort showed significant stratification, outperforming manual cell counting. Visual comparison between the two algorithms revealed that the deep learning algorithm produced more faithful segmentation of the nuclei and was more adaptive to variation in cell size that is inherent to different stages and differentiation grades of CRC, thus the measures of agreement favoured utilising the deep learning algorithm in subsequent cohorts. The use of image analysis in fundamental research is generally coupled with some level of manual histopathological assessment to ensure reproducibility, however, in a clinical setting where the aim of using image analysis is to produce prognostically significant data without requiring additional time from the diagnosing pathologist. Hence any image analysis

algorithm that would be implemented in a clinical setting must be able to reliably adapt to novel image data and still produce accurate results.

The GRI cohort is a newer patient cohort than the two cohorts used to validate the image analysis and had not been previously scored for Ki67 expression. Therefore, to simulate clinical application of this image analysis approach, the algorithm was used to assess Ki67 expression in the GRI cohort without adaptation to any histological processing variation, such as stain intensity, in this cohort and patients were stratified for high vs low expression using the same </> 30% PI cut off value that had been established in prior analyses. The deep learning algorithm significantly stratified these patients for CSS, confirming high Ki67 expression as a favourable prognostic factor and the accuracy of the </> 30% PI cut off value, and further demonstrating the ability of deep learning – based image analysis to translate to novel patient data. It is highly likely that the first image analysis algorithms to see clinical use will be for conducting relatively straightforward although time consuming assessments such as this, and in the setting of breast cancer, image analysis is currently under validation for assessing HER2 expression to distinguish molecular subtypes, thus this study adds to the growing evidence of the clinical applications of image analysis.

The translational capacity of this approach was further confirmed in the large TransSCOT clinical trial cohort. A TMA was constructed for this cohort with the intention that it be used for translational studies, and as such, one core from the tumour core and one core from the invasive margin was included for each patient. It is understood that Ki67 expression is variable between tumour subsites, and within the epithelial crypts depending on disease state, however the prognostic effect of this has yet to be fully elucidated, with studies producing conflicting results. To investigate further the effect of subsite dependent Ki67 expression, the deep learning algorithm was applied to both the tumour core and invasive margin cores for each patient in the TMA. Stratifying patients for high vs low expression using the </> 30% PI cut off revealed both sites to be prognostically significant for DFS, with Ki67 expression appearing to be more prognostic at the invasive margin. It is of note however, that the distribution of patients between high and low expression is more even at the invasive margin, and skewed towards high expression in the tumour core, which is likely affecting the measure of statistical significance. It is likely that further investigation of this phenomenon using full sections will elucidate the true effect of site dependent Ki67 expression on prognosis. Ki67 expression is also known to vary between primary tumour site within the bowel although the effect this has on prognosis has also yet to be firmly established. Data from this cohort demonstrates that expression in the tumour core is

significantly higher in both left and right sided disease, and while there is no site dependent variation in expression at the invasive margin, there is a non – significant increase in tumour centre expression in right sided disease. This could be due to the higher proportion of right sided tumours that are MSI, which are known to be associated with a *KRAS* mediated increase in proliferative activity. Investigating the possible affect this may have on survival revealed that there is a greater risk associated with low Ki67 expression in the tumour core than at the invasive margin in left sided disease, but the risk is identical in right sided disease, suggesting that Ki67 expression may need to be assessed differently depending on disease site.

These data further demonstrate that image analysis can be reliably used to conduct fundamental research in an unbiased manner without requiring additional human input to adapt the algorithms to novel data or counter score a subset of each new patient cohort to ensure reliability. This is of course dependent upon the task that the image analysis is set to. Positive cell counting has been conducted using image analysis almost since its inception and is therefore extremely reliable for certain markers such as Ki67, which is a well – defined nuclear stain produced using thoroughly validated diagnostic antibodies. However, more experimental biomarkers with variable expression patterns or markers which localise to cellular compartments that often prove difficult to accurately score, such as the membrane or nucleolus, would require rigorous validation through image analysis and possibly visual confirmation of classification accuracy by a diagnosing pathologist for each patient being assessed.

In Chapter 6, the three prognostic features assessed in Chapters 3 – 5 were combined into established classification systems and assessed for prognostic significance against the manual histopathological counterparts. The GMS is a combinatorial score of the KM grade and TSP, therefore, to grade patients using these criteria though image analysis, the TSP algorithm, semi – automated lymphocyte quantification algorithm, and automated lymphocyte quantification algorithm were combined on a single H&E per patient. In the Glasgow Development, Glasgow Validation, and TransSCOT cohorts, GMS stratification using the semi – automated lymphocyte quantification and TSP algorithms significantly stratified patients for CSS and DFS following the same pattern as is found through manual histopathological assessment, with GMS0 possessing the best prognosis, GMS2 the worst prognosis, and GMS1 intermediate prognosis. Additionally, the image analysis approach outperformed manual histopathological assessment in the Glasgow Development cohort and near identically in the TransSCOT cohort. Previous studies have shown that combinations

of lymphocyte subset quantification through IHC and TSP assessment hold prognostic significance through image analysis, however, to the authors knowledge, a combination score of these metrics has not been performed on H&E WSI in large patient cohorts. While this assessment requires prior pathologist input in the form of annotation of the invasive margin, data from Chapter 4 shows this to be reliable between observers and does not require sufficient time to be prohibitive to use in a clinical setting, considering the significant association of this grading system with prognosis. Furthermore, automated assessment of the GMS through application of the lymphocyte detection algorithm to the stromal areas identified by the TSP algorithm retained a significant association with survival in all three cohorts in this study. It is therefore possible that with an improved lymphocyte detection algorithm, such as those previously discussed, that is able to identify specific lymphocyte populations through deep learning, automated assessment could provide equivalent prognostic information without the associated time requirement. This would be contingent on further investigation of the difference in prognostic and clinical significance of peritumoural vs intratumoural immune response.

The algorithms used to produce the GMS classification system were further combined with the Ki67 assessment detailed in Chapter 5 to determine the Phenotypic Subtype of each patient through image analysis. Using this criteria, GMS1 patients with intermediate prognosis are stratified into good and poor prognostic groups dependent upon their Ki67 %PI, thus GMS0 / Immune subtype patients retain the best prognosis, GMS2 / Stromal subtype patients retain the worst prognosis, and high Ki67 expression GMS1 / Canonical subtype patients possess an intermediate good prognosis with low for all metrics GMS1 / Latent subtype patients possessing an intermediate poor prognosis. Phenotypic Subtypes determined through image analysis with semi – automated lymphocyte detection showed significant associations with survival in all three cohorts used in this study, as did Phenotypic Subtypes determined through automated image analysis. Although the disproportionate number of Immune subtype patients determined through automated image analysis resulted in Canonical subtype patients possessing the best prognosis in the Glasgow Validation cohort and a weakened association with survival in the Glasgow Development and TransSCOT cohorts. As described in Chapter 6, Roseweir and Colleagues derived the Phenotypic Subtypes from the phenotypic signatures of the individual CMS subtypes, and while previous studies have investigated the translation of the CMS to image analysis, the subtyping system did not retain association with survival (Roseweir et al., 2017; Sirinukunwattana, Domingo, Richman, Redmond, Blake, Verrill, Leedham, Chatzipli, Hardy, Whalley, Wu, Beggs, McDermott, Dunne, Meade, Walker, Murray, Samuel,

Seymour, Tomlinson, Quirke, Maughan, Rittscher, & Koelzer, 2021). This study therefore can serve as a framework for interpreting genomics data to image analysis, whereby pathologists interpret the morphological characteristics of subtypes derived from genomic analyses and image analysis approaches are developed to analyse these characteristics. This study demonstrates that through this method, the prognostic characteristics of specific genomic alterations can be carried forward to image analysis. Previous studies have demonstrated the utility of this kind of investigation through using image analysis to assess the unique histological features of the tumour associated with MSI status (Nguyen et al., 2021).

Further support to the clinical applicability of the image analysis algorithms was demonstrated through assessment of chemotherapy response. During validation of the Phenotypic Subtypes in the TransSCOT cohort, Roseweir and Colleagues identified that Immune subtype patients possessed a significantly improved prognosis when administered FOLFOX chemotherapy over CAPOX. It is imperative to the clinical validation of image analysis algorithms that they are able to identify the same biological phenomena as manual histopathological, particularly pertaining the theragnosis, as this would be of substantial utility to physicians when directing patient care. Determining the Phenotypic Subtypes using the semi – automated lymphocyte detection algorithm, Immune subtype patients showed an increased risk of poor prognosis when administered CAPOX over FOLFOX. While this difference was not significant, the hazard ratio for patients receiving CAPOX was 1.6 and no other difference in survival between chemotherapy regimens was seen in the other subtypes. It is possible that further investigation into the lymphocyte response at the invasive margin and development of a method by which to faithfully assess the KM grading criteria through image analysis, could elucidate the underlying mechanisms of improved chemotherapy – dependent prognosis and indeed, provide a means to reliably identify patients that could benefit from different chemotherapies. These data demonstrate that the concerted application of image analysis algorithms is able to accurately stratify patients according to novel classification criteria and potentially influence chemotherapeutic decisions to benefit patients.

In summary, this thesis has demonstrated that image analysis can be successfully applied to novel prognostic features assessed on clinical H&E WSI, be reliably translated between independent patient cohorts, and reproducibly stratify patients for prognosis to a greater significance than manual histopathological assessment for certain criteria. Furthermore, combinations of image analysis algorithms can be applied to novel, multi –

stage classification criteria while retaining prognostic significance, with the potential to identify patient subsets that would benefit from targeted chemotherapy.

7.2 Future Work

This work outlines a framework for developing and validating image analysis for potential clinical use, as well as investigating preliminary image analysis algorithms to conduct novel assessments. Work following on from this thesis should be centred around ensuring the clinical translatability of the image analysis approaches to the novel prognostic features.

The TSP algorithm performs well for patient stratification however, as shown in Chapter 3 the optimal cut off value for survival is 40%, markedly different from the 50% determined via manual assessment. It is likely that this is due to the fact that only 50% of the tumour area is analysed with the current algorithm due to computational limitations. This could be addressed as computing technology develops and is able to cope with whole section semantic segmentation via deep learning. In the immediate future however, Ho and Colleagues demonstrated that instigating a neural network training protocol operating at multiple objective magnifications is able to perform semantic segmentation of breast cancer WSI at lower magnifications while retaining classification accuracy (Ho et al., 2021). A similar methodology could be applied to the TSP assessment in CRC using the algorithm implemented in this study, allowing the whole tumour area to be classified thus improving the reliability of the image analysis assessment.

As previously stated, the granular prognostic detail of the KM grading system is likely a result of the morphological nuances of the assessment, something which is evidently not recapitulated through lymphocyte density image analysis. There is a rich body of work detailing the spatial relationships between tumour infiltrating lymphocytes, neighbouring immune cell populations, the tumour, and other features of the CRC TME that can influence prognosis (Corredor et al., 2019; Elomaa et al., 2022; M. Jung et al., 2022; I. Nearchou et al., 2021; I. P. Nearchou, Lillard, et al., 2019c; Xu et al., 2022). Examining the spatial relationships of the TME cell types in the context of each KM grade would likely elucidate a means by which to more faithfully adapt the KM criteria to image analysis. Additionally, the distinct pattern of inflammatory response of each KM grade could also provide a means by which to adapt the criteria to deep learning – based image analysis, as patterns of lymphocytic infiltrate have previously been shown to modulate prognosis and response to chemo therapeutics in a variety of malignancies (Lu et al., 2020; Saltz et al., 2018; Shaban et al., 2019; X. Wang et al., 2022; B. Zhang et al., 2021). While a deep learning – based

approach such as this would be a more “black box” – like method than what was utilised in this study, it is plausible that a combination of a deep learning generated probability of the KM grade coupled with more transparent, traditional image analysis approaches would provide additional prognostic information and strengthen the reliability of the assessment.

IHC is a routine part of diagnostic pathology, providing molecular information about the individual tumour and often guiding subsequent clinical decisions. While IHC is a relatively cost – effective method, any reduction in associated cost will contribute to the clinical utility of a novel methodology. It would therefore be of interest to elucidate a means by which to infer, or directly measure, the proliferative activity of a tumour from a single H&E slide without the need for a Ki67 IHC slide to be produced. Not only would this reduce the cost of determining the phenotypic subtypes, but it would also save precious tissue from the diagnostic block. A direct method of measuring tumour proliferation on H&E sections is through mitotic figure indexing, which is already a prognostic assessment in breast cancer pathology. Both traditional image analysis and deep learning approaches have showed excellent efficacy in detecting mitotic figures on H&E sections, subsequently relating them to chemotherapy response and prognosis (Balkenhol et al., 2019; Bigley et al., 2016; Romo-Bucheli et al., 2017). However, it is an established principal that Ki67 indices and mitotic figure counts do not correlate well, if at all, therefore a study validating the efficacy of mitotic figures to stratify GMS1 patients into the Canonical and Latent subtypes would be necessary. An indirect method by which Ki67 positivity status could be measured on H&E slides is through antibody supervised deep learning, where IHC sections stained for specific markers are used as ground truth for neural network training which is then translated to H&E sections. This method has been proven to be effective for distinguishing lymphocyte populations and accurately segmenting epithelial tissue following training on pan – cytokeratin (CK) IHC sections (Bulten et al., 2019; Stenman et al., 2021; Turkki et al., 2016), and it is therefore plausible that Ki67 IHC guided deep learning could discern high dimensional features to detect cells likely to be $Ki67^{+}$ on H&E slides.

8. References

Acs, Balazs, Pelekanou, Vasiliki, Bai, Yalai, Martinez-Morilla, Sandra, Toki, Maria, Leung, Samuel C. Y., Nielsen, Torsten O., Rimm, David L. (2018). Ki67 reproducibility using digital image analysis: an inter-platform and inter-operator study. *Laboratory Investigation 2018 99:1*, *99*(1), 107–117.

Advani, Shailesh M., Advani, Pragati, DeSantis, Stacia M., Brown, Derek, VonVille, Helena M., Lam, Michael, Loree, Jonathan M., Sarshekeh, Amir Mehrvarz, Bressler, Jan, Lopez, David S., Daniel, Carrie R., Swartz, Michael D., Kopetz, Scott. (2018). Clinical, Pathological, and Molecular Characteristics of CpG Island Methylator Phenotype in Colorectal Cancer: A Systematic Review and Meta-analysis. *Translational Oncology*, *11*(5), 1188–1201.

Aizawa, Takashi, Karasawa, Hideaki, Funayama, Ryo, Shirota, Matsuyuki, Suzuki, Takashi, Maeda, Shimpei, Suzuki, Hideyuki, Yamamura, Akihiro, Naitoh, Takeshi, Nakayama, Keiko, Unno, Michiaki. (2019). Cancer-associated fibroblasts secrete Wnt2 to promote cancer progression in colorectal cancer. *Cancer Medicine*, *8*(14), 6370–6382.

Alexander, Peter G., McMillan, Donald C., Park, James H. (2020). The local inflammatory response in colorectal cancer – Type, location or density? A systematic review and meta-analysis. *Cancer Treatment Reviews*, *83*, 101949.

Alexander, Peter G., Roseweir, Antonia K., Pennel, Kathryn A. F., van Wyk, Hester C., Powell, Arfon G. M. T., McMillan, Donald C., Horgan, Paul G., Kelly, Caroline, Hay, Jennifer, Sansom, Owen, Harkin, Andrea, Roxburgh, Campbell S. D., Graham, Janet, Church, David N., Tomlinson, Ian, Saunders, Mark, Iveson, Tim J., Edwards, Joanne, Park, James H. (2021a). The Glasgow Microenvironment Score associates with prognosis and adjuvant chemotherapy response in colorectal cancer. *British Journal of Cancer*, *124*(4), 786–796.

Alexander, Peter G., Roseweir, Antonia K., Pennel, Kathryn A. F., van Wyk, Hester C., Powell, Arfon G. M. T., McMillan, Donald C., Horgan, Paul G., Kelly, Caroline, Hay, Jennifer, Sansom, Owen, Harkin, Andrea, Roxburgh, Campbell S. D., Graham, Janet, Church, David N., Tomlinson, Ian, Saunders, Mark, Iveson, Tim J., Edwards, Joanne, Park, James H. (2021b). The Glasgow Microenvironment Score associates with prognosis and adjuvant chemotherapy response in colorectal cancer. *British Journal of Cancer*, *124*(4), 786–796.

Almangush, Alhadi, Pirinen, Matti, Heikkinen, Ilkka, Mäkitie, Antti A., Salo, Tuula, Leivo, Ilmo. (2017). Tumour budding in oral squamous cell carcinoma: a meta-analysis. *British Journal of Cancer 2018 118:4*, *118*(4), 577–586.

Ao, Tadakazu, Kajiwara, Yoshiki, Yonemura, Keisuke, Shinto, Eiji, Mochizuki, Satsuki, Okamoto, Koichi, Aosasa, Suefumi, Ueno, Hideki. (2019). Prognostic significance of histological categorization of desmoplastic reaction in colorectal liver metastases. *Virchows Archiv*, *475*(3), 341–348.

Ao, Tadakazu, Kajiwara, Yoshiki, Yonemura, Keisuke, Shinto, Eiji, Mochizuki, Satsuki, Okamoto, Koichi, Kishi, Yoji, Ueno, Hideki. (2020). Morphological consistency of desmoplastic reactions between the primary colorectal cancer lesion and associated metastatic lesions. *Virchows Archiv*, *477*(1), 47–55.

Ao, Tadakazu, Mochizuki, Satsuki, Kajiwara, Yoshiki, Yonemura, Keisuke, Shiraishi, Takehiro, Nagata, Ken, Shinto, Eiji, Okamoto, Koichi, Nearchou, Ines P., Shimazaki, Hideyuki, Kishi, Yoji, Okada, Yasunori, Ueno, Hideki. (2022). Cancer-associated fibroblasts at the unfavorable desmoplastic stroma promote colorectal cancer aggressiveness: Potential role of ADAM9. *International Journal of Cancer*, *150*(10), 1706–1721.

Armaghany, Tannaz, Wilson, Jon D., Chu, Quyen, Mills, Glenn. (2012). Genetic Alterations in Colorectal Cancer. *Gastrointestinal Cancer Research : GCR*, *5*(1), 19.

Artificial Intelligence to help save lives at five new technology centres - GOV.UK. (n.d.). Retrieved 26 May 2023, from

Arvaniti, Eirini, Fricker, Kim S., Moret, Michael, Rupp, Niels, Hermanns, Thomas, Fankhauser, Christian, Wey, Norbert, Wild, Peter J., Rüschoff, Jan H., Claassen, Manfred. (2018). Automated Gleason grading of prostate cancer tissue microarrays via deep learning. *Scientific Reports 2018 8:1*, *8*(1), 1–11.

Balkenhol, Maschenka C. A., Tellez, David, Vreuls, Willem, Clahsen, Pieter C., Pinckaers, Hans, Ciompi, Francesco, Bult, Peter, van der Laak, Jeroen A. W. M. (2019). Deep learning assisted mitotic counting for breast cancer. *Laboratory Investigation 2019 99:11*, *99*(11), 1596–1606.

Bamford, Pascal, Lovell, Brian. (2001). Method for accurate unsupervised cell nucleus segmentation. *Annual Reports of the Research Reactor Institute, Kyoto University*, *3*, 2704–2708.

Bándi, Péter, Geessink, Oscar, Manson, Quirine, Van Dijk, Marcory, Balkenhol, Maschenka, Hermsen, Meyke, Ehteshami Bejnordi, Babak, Lee, Byungjae, Paeng, Kyunghyun, Zhong, Aoxiao, Li, Quanzheng, Zanjani, Farhad Ghazvinian, Zinger, Svitlana, Fukuta, Keisuke, Komura, Daisuke, Ovtcharov, Vlado, Cheng, Shenghua, Zeng, Shaoqun, Thagaard, Jeppe, Dahl, Anders B., Lin, Huangjing, Chen, Hao, Jacobsson, Ludwig, Hedlund, Martin, Çetin, Melih, Halici, Eren, Jackson, Hunter, Chen, Richard, Both, Fabian, Franke, Jörg, Kusters-Vandevelde, Heidi, Vreuls, Willem, Bult, Peter, Van Ginneken, Bram, Van Der Laak, Jeroen, Litjens, Geert. (2019). From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge. *IEEE Transactions on Medical Imaging*, *38*(2), 550–560.

Bankhead, Peter. (2022). Developing image analysis methods for digital pathology. *The Journal of Pathology*, *257*(4), 391–402.

Bankhead, Peter, Loughrey, Maurice B., Fernández, José A., Dombrowski, Yvonne, McArt, Darragh G., Dunne, Philip D., McQuaid, Stephen, Gray, Ronan T., Murray, Liam J., Coleman, Helen G., James, Jacqueline A., Salto-Tellez, Manuel, Hamilton, Peter W. (2017). QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, *7*(1).

Bartley, Angela N., Luthra, Rajyalakshmi, Saraiya, Devki S., Urbauer, Diana L., Broaddus, Russell R. (2012). Identification of cancer patients with lynch syndrome: Clinically significant discordances and problems in tissue-based mismatch repair testing. *Cancer Prevention Research*, *5*(2), 320–327.

Bateman, Adrian C. (2021). DNA mismatch repair proteins: scientific update and practical guide. *Journal of Clinical Pathology*, *74*(4), 264–268.

Becht, Etienne, De Reyniès, Aurélien, Giraldo, Nicolas A., Pilati, Camilla, Buttard, Bénédicte, Lacroix, Laetitia, Selves, Janick, Sautès-Fridman, Catherine, Laurent-Puig, Pierre, Fridman, Wolf Herman. (2016a). Immune and stromal classification of Colorectal cancer is associated with molecular subtypes and relevant for precision immunotherapy. *Clinical Cancer Research*, *22*(16), 4057–4066.

Becht, Etienne, De Reyniès, Aurélien, Giraldo, Nicolas A., Pilati, Camilla, Buttard, Bénédicte, Lacroix, Laetitia, Selves, Janick, Sautès-Fridman, Catherine, Laurent-Puig, Pierre, Fridman, Wolf Herman. (2016b). Immune and stromal classification of Colorectal cancer is associated with molecular subtypes and relevant for precision immunotherapy. *Clinical Cancer Research*, *22*(16), 4057–4066.

Bejnordi, Babak Ehteshami, Litjens, Geert, Timofeeva, Nadya, Otte-Höller, Irene, Homeyer, André, Karssemeijer, Nico, Van Der Laak, Jeroen A. W. M. (2016). Stain specific standardization of whole-slide histopathological images. *IEEE Transactions on Medical Imaging*, *35*(2), 404–415.

Bejnordi, Babak Ehteshami, Veta, Mitko, Van Diest, Paul Johannes, Van Ginneken, Bram, Karssemeijer, Nico, Litjens, Geert, Van Der Laak, Jeroen A. W. M., Hermsen, Meyke, Manson, Quirine F., Balkenhol, Maschenka, Geessink, Oscar, Stathonikos, Nikolaos, Van Dijk, Marcory C. R. F., Bult, Peter, Beca, Francisco, Beck, Andrew H., Wang, Dayong, Khosla, Aditya, Gargeya, Rishab, Irshad, Humayun, Zhong, Aoxiao, Dou, Qi, … Venâncio, Rui. (2017). Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*, *318*(22), 2199–2210.

Belli, Sedat, Aytac, Huseyin Ozgur, Karagulle, Erdal, Yabanoglu, Hakan, Kayaselcuk, Fazilet, Yildirim, Sedat. (2014). Outcomes of Surgical Treatment of Primary Signet Ring Cell Carcinoma of the Colon and Rectum: 22 Cases Reviewed With Literature. *International Surgery*, *99*(6), 691.

Bellis, D., Marci, V., Monga, G. (1993). Light Microscopic and Immunohistochemical Evaluation of Vascular and Neural Invasion in Colorectal Cancer. *Pathology - Research and Practice*, *189*(4), 443–447.

Benatti, Piero, Gafà, Roberta, Barana, Daniela, Marino, Massimiliano, Scarselli, Alessandra, Pedroni, Monica, Maestri, Iva, Guerzoni, Laura, Roncucci, Luca, Menigatti, Mirco, Roncari, Barbara, Maffei, Stefania, Rossi, Giuseppina, Ponti, Giovanni, Santini, Alessandra, Losi, Lorena, Di Gregorio, Carmela, Oliani, Cristina, De Leon, Maurizio Ponz, Lanza, Giovanni. (2005). Microsatellite Instability and Colorectal Cancer Prognosis. *Clinical Cancer Research*, *11*(23), 8332–8340.

Berbís, M. Alvaro, McClintock, David S., Bychkov, Andrey, Van der Laak, Jeroen, Pantanowitz, Liron, Lennerz, Jochen K., Cheng, Jerome Y., Delahunt, Brett, Egevad, Lars, Eloy, Catarina, Farris, Alton B., Fraggetta, Filippo, García del Moral, Raimundo, Hartman, Douglas J., Herrmann, Markus D., Hollemans, Eva, Iczkowski, Kenneth A., Karsan, Aly, Kriegsmann, Mark, Salama, Mohamed E., Sinard, John H., Tuthill, J. Mark, Williams, Bethany, Casado-Sánchez, César, Sánchez-Turrión, Víctor, Luna, Antonio, Aneiros-Fernández, José, Shen, Jeanne. (2023). Computational pathology in

2030: a Delphi study forecasting the role of AI in pathology within the next decade. *EBioMedicine*, *88*, 104427.

Berg, Alfred O., Armstrong, Katrina, Botkin, Jeffrey, Calonge, Ned, Haddow, James, Hayes, Maxine, Kaye, Celia, Phillips, Kathryn A., Piper, Margaret, Richards, Carolyn Sue, Scott, Joan A., Strickland, Ora L., Teutsch, Steven. (2009). Recommendations from the EGAPP Working Group: genetic testing strategies in newly diagnosed individuals with colorectal cancer aimed at reducing morbidity and mortality from Lynch syndrome in relatives. *Genetics in Medicine*, *11*(1), 35–41.

Béroud, Christophe, Soussi, Thierry. (1996). APC gene: database of germline and somatic mutations in human tumors and cell lines. *Nucleic Acids Research*, *24*(1), 121–124.

Betge, Johannes, Pollheimer, Marion J., Lindtner, Richard A., Kornprat, Peter, Schlemmer, Andrea, Rehak, Peter, Vieth, Michael, Hoefler, Gerald, Langner, Cord. (2012). Intramural and extramural vascular invasion in colorectal cancer: prognostic significance and quality of pathology reporting. *Cancer*, *118*(3), 628–638.

Bigley, Alison L., Klein, Stephanie K., Davies, Barry, Williams, Leigh, Rudmann, Daniel G. (2016). Using Automated Image Analysis Algorithms to Distinguish Normal, Aberrant, and Degenerate Mitotic Figures Induced by Eg5 Inhibition. *Toxicologic Pathology*, *44*(5), 663–672.

Bird, Adrian P. (1986). CpG-rich islands and the function of DNA methylation. *Nature 1986 321:6067*, *321*(6067), 209–213.

Bohaumilitzky, Lena, Huber, Ann Kathrin, Stork, Eva Maria, Wengert, Simon, Woelfl, Franziska, Boehm, Heike. (2017). A trickster in disguise: Hyaluronan's ambivalent roles in the matrix. *Frontiers in Oncology*, *7*(OCT), 242.

Bosch, Stevenl, Teerenstra, Steven, De Wilt, Johannesh W., Cunningham, Chris, Nagtegaal, Irisd. (2013). Predicting lymph node metastasis in pT1 colorectal cancer: a systematic review of risk factors providing rationale for therapy decisions. *Endoscopy*, *45*(10), 827–834.

Bosman, F. T., Carneiro, F., Hruban, R. H., Theise, N. D. (2010a). WHO classification of tumours of the digestive system. *WHO Classification of Tumours of the Digestive System.*, *Ed. 4*.

Bosman, F. T., Carneiro, F., Hruban, R. H., Theise, N. D. (2010b). WHO Classification of Tumours of the Digestive System. In *WHO press*.

Bouzubar, N., Walker, K. J., Griffiths, K., Ellis, I. O., Elston, C. W., Robertson, J. F. R., Blamey, R. W., Nicholson, R. I. (1989). Ki67 immunostaining in primary breast cancer: pathological and clinical associations. *British Journal of Cancer 1989 59:6*, *59*(6), 943–947.

Brabletz, T., Jung, A., Reu, S., Porzner, M., Hlubek, F., Kunz-Schughart, L. A., Knuechel, R., Kirchner, T. (2001). Variable β-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(18), 10356–10361.

Browning, Lisa, Fryer, Eve, Roskell, Derek, White, Kieron, Colling, Richard, Rittscher, Jens, Verrill, Clare. (2021). Role of digital pathology in diagnostic histopathology in the response to COVID-19: results from a survey of experience in a UK tertiary referral hospital. *Journal of Clinical Pathology*, *74*(2), 129–132.

Brügmann, Anja, Eld, Mikkel, Lelkaitis, Giedrius, Nielsen, Søren, Grunkin, Michael, Hansen, Johan D., Foged, Niels T., Vyberg, Mogens. (2012). Digital image analysis of membrane connectivity is a robust measure of HER2 immunostains. *Breast Cancer Research and Treatment*, *132*(1), 41–49.

Budinska, Eva, Popovici, Vlad, Tejpar, Sabine, D'Ario, Giovanni, Lapique, Nicolas, Sikora, Katarzyna Otylia, Di Narzo, Antonio Fabio, Yan, Pu, Graeme Hodgson, John, Weinrich, Scott, Bosman, Fred, Roth, Arnaud, Delorenzi, Mauro. (2013). Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *The Journal of Pathology*, *231*(1), 63–76.

Bufill, J. A. (1990). Colorectal cancer: Evidence for distinct genetic categories based on proximal or distal tumor location. *Annals of Internal Medicine*, *113*(10), 779–788.

Bui, Marilyn M., Riben, Michael W., Allison, Kimberly H., Chlipala, Elizabeth, Colasacco, Carol, Kahn, Andrea G., Lacchetti, Christina, Madabhushi, Anant, Pantanowitz, Liron, Salama, Mohamed E., Stewart, Rachel L., Thomas, Nicole E., Tomaszewski, John E., Hammond, M. Elizabeth. (2019). Quantitative Image Analysis of Human Epidermal Growth Factor Receptor 2 Immunohistochemistry for Breast Cancer: Guideline From the College of American Pathologists. *Archives of Pathology & Laboratory Medicine*, *143*(10), 1180.

Bulten, Wouter, Bándi, Péter, Hoven, Jeffrey, Loo, Rob van de, Lotz, Johannes, Weiss, Nick, Laak, Jeroen van der, Ginneken, Bram van, Hulsbergen-van de Kaa, Christina, Litjens,

Geert. (2019). Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard. *Scientific Reports 2019 9:1*, *9*(1), 1–10.

Burt, Randall W., Leppert, Mark F., Slattery, Martha L., Samowitz, Wade S., Spirio, Lisa N., Kerber, Richard A., Kuwada, Scott K., Neklason, Deborah W., Disario, James A., Lyon, Elaine, Hughes, J. Preston, Chey, William Y., White, Raymond L. (2004). Genetic testing and phenotype in a large kindred with attenuated familial adenomatous polyposis. *Gastroenterology*, *127*(2), 444–451.

Bussey, HJ. (1975). Familial polyposis coli. *Family Studies, Histopathology, Differential Diagnosis, and Results of Treatment*, 47–49.

Bychkov, Dmitrii, Linder, Nina, Turkki, Riku, Nordling, Stig, Kovanen, Panu E., Verrill, Clare, Walliander, Margarita, Lundin, Mikael, Haglund, Caj, Lundin, Johan. (2018a). Deep learning based tissue analysis predicts outcome in colorectal cancer. *Scientific Reports 2018 8:1*, *8*(1), 1–11.

Bychkov, Dmitrii, Linder, Nina, Turkki, Riku, Nordling, Stig, Kovanen, Panu E., Verrill, Clare, Walliander, Margarita, Lundin, Mikael, Haglund, Caj, Lundin, Johan. (2018b). Deep learning based tissue analysis predicts outcome in colorectal cancer. *Scientific Reports*, *8*(1), 3395.

Bylsma, Lauren C., Gillezeau, Christina, Garawin, Tamer A., Kelsh, Michael A., Fryzek, Jon P., Sangaré, Laura, Lowe, Kimberly A. (2020). Prevalence of RAS and BRAF mutations in metastatic colorectal cancer patients by tumor sidedness: A systematic review and meta-analysis. *Cancer Medicine*, *9*(3), 1044–1057.

Calon, Alexandre, Espinet, Elisa, Palomo-Ponce, Sergio, Tauriello, Daniele V. F., Iglesias, Mar, Céspedes, María Virtudes, Sevillano, Marta, Nadal, Cristina, Jung, Peter, Zhang, Xiang H. F., Byrom, Daniel, Riera, Antoni, Rossell, David, Mangues, Ramón, Massagué, Joan, Sancho, Elena, Batlle, Eduard. (2012a). Dependency of Colorectal Cancer on a TGF-β-Driven Program in Stromal Cells for Metastasis Initiation. *Cancer Cell*, *22*(5), 571–584.

Calon, Alexandre, Espinet, Elisa, Palomo-Ponce, Sergio, Tauriello, Daniele V. F., Iglesias, Mar, Céspedes, María Virtudes, Sevillano, Marta, Nadal, Cristina, Jung, Peter, Zhang, Xiang H. F., Byrom, Daniel, Riera, Antoni, Rossell, David, Mangues, Ramón, Massagué, Joan, Sancho, Elena, Batlle, Eduard. (2012b). Dependency of Colorectal

Cancer on a TGF-β-Driven Program in Stromal Cells for Metastasis Initiation. *Cancer Cell*, *22*(5), 571–584.

Calon, Alexandre, Lonardo, Enza, Berenguer-Llergo, Antonio, Espinet, Elisa, Hernando-Momblona, Xavier, Iglesias, Mar, Sevillano, Marta, Palomo-Ponce, Sergio, Tauriello, Daniele V. F., Byrom, Daniel, Cortina, Carme, Morral, Clara, Barceló, Carles, Tosi, Sebastien, Riera, Antoni, Attolini, Camille Stephan Otto, Rossell, David, Sancho, Elena, Batlle, Eduard. (2015). Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nature Genetics 2015 47:4*, *47*(4), 320–329.

Camp, Robert L., Chung, Gina G., Rimm, David L. (2002). Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nature Medicine 2002 8:11*, *8*(11), 1323–1328.

Campanella, Gabriele, Hanna, Matthew G., Geneslaw, Luke, Miraflor, Allen, Werneck Krauss Silva, Vitor, Busam, Klaus J., Brogi, Edi, Reuter, Victor E., Klimstra, David S., Fuchs, Thomas J. (2019a). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine 2019 25:8*, *25*(8), 1301–1309.

Campanella, Gabriele, Hanna, Matthew G., Geneslaw, Luke, Miraflor, Allen, Werneck Krauss Silva, Vitor, Busam, Klaus J., Brogi, Edi, Reuter, Victor E., Klimstra, David S., Fuchs, Thomas J. (2019b). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, *25*(8), 1301–1309.

Cappellesso, Rocco, Luchini, Claudio, Veronese, Nicola, Lo Mele, Marcello, Rosa-Rizzotto, Erik, Guido, Ennio, De Lazzari, Franca, Pilati, Pierluigi, Farinati, Fabio, Realdon, Stefano, Solmi, Marco, Fassan, Matteo, Rugge, Massimo. (2017). Tumor budding as a risk factor for nodal metastasis in pT1 colorectal cancers: a meta-analysis. *Human Pathology*, *65*, 62–70.

Capper, David, Voigt, Anita, Bozukova, Gergana, Ahadova, Aysel, Kickingereder, Philipp, Von Deimling, Andreas, Von Knebel Doeberitz, Magnus, Kloor, Matthias. (2013). BRAF V600E-specific immunohistochemistry for the exclusion of Lynch syndrome in MSI-H colorectal cancer. *International Journal of Cancer*, *133*(7), 1624–1630.

Caul, Sarah, Broggio, John. (n.d.). *Cancer registration statistics, England Statistical bulletins - Office for National Statistics*. Retrieved 24 August 2022, from

Chan, Annie On On, Issa, Jean Pierre J., Morris, Jeffrey S., Hamilton, Stanley R., Rashid, Asif. (2002). Concordant CpG Island Methylation in Hyperplastic Polyposis. *The American Journal of Pathology*, *160*(2), 529–536.

Che, Yun, Wang, Jingnan, Li, Yuan, Lu, Zhiliang, Huang, Jianbing, Sun, Shouguo, Mao, Shuangshuang, Lei, Yuanyuan, Zang, Ruochuan, Sun, Nan, He, Jie. (2018). Cisplatin-activated PAI-1 secretion in the cancer-associated fibroblasts with paracrine effects promoting esophageal squamous cell carcinoma progression and causing chemoresistance. *Cell Death & Disease*, *9*(7).

Chen, Liang Chieh, Zhu, Yukun, Papandreou, George, Schroff, Florian, Adam, Hartwig. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11211 LNCS*, 833–851.

Cheng, Lee, Eng, Cathy, Nieman, Linda Z., Kapadia, Asha S., Du, Xianglin L. (2011). Trends in colorectal cancer incidence by anatomic site and disease stage in the United States from 1976 to 2005. *American Journal of Clinical Oncology: Cancer Clinical Trials*, *34*(6), 573–580.

Chollet, François. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *2017-January*, 1800–1807.

Christgen, Matthias, Von Ahsen, Sabrina, Christgen, Henriette, Länger, Florian, Kreipe, Hans. (2015). The region-of-interest size impacts on Ki67 quantification by computer-assisted image analysis in breast cancer. *Human Pathology*, *46*(9), 1341–1349.

Clark, Gavin Rc, Anderson, Annie S., Godfrey, Thomas G., Strachan, Judith A., Fraser, Callum G., Steele, Robert Jc. (2020). Variation in changes in the incidence of colorectal cancer by age and association with screening uptake: an observational study. *BMJ Open*, *10*, 37925.

Colling, Richard, Pitman, Helen, Oien, Karin, Rajpoot, Nasir, Macklin, Philip, Bachtiar, Velicia, Booth, Richard, Bryant, Alyson, Bull, Joshua, Bury, Jonathan, Carragher, Fiona, Collins, Graeme, Craig, Clare, da Silva, Maria Freitas, Gosling, Daniel, Jacobs, Jaco, Kajland-Wilén, Lena, Karling, Johanna, Lawler, Darragh, Lee, Stephen, Miller, Keith, Mozolowski, Guy, Nicholson, Richard, O'Connor, Daniel, Rahbek, Mikkel, Sumner, Alan, Vossen, Dirk, White, Kieron, Wing, Charlotte, Wright, Corrina, Snead,

David, Sackville, Tony, Verrill, Clare. (2019). Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. *The Journal of Pathology*, *249*(2), 143–150.

Conte, Gabriella A., Qari, Omar, Fasano, Genevieve A., Guinto, Robyn K., Palo, Laura, Parker, Glenn S., Rangwala, Anis F., Minassian, Haig, Greenberg, Patricia J., Dewan, Asa A., Topilow, Arthur A. (2020). S100 Staining Adds to the Prognostic Significance of the Combination of Perineural Invasion and Lymphovascular Invasion in Colorectal Cancer. *Applied Immunohistochemistry & Molecular Morphology : AIMM*, *28*(5).

Corredor, German, Wang, Xiangxue, Zhou, Yu, Lu, Cheng, Fu, Pingfu, Syrigos, Konstantinos, Rimm, David L., Yang, Michael, Romero, Eduardo, Schalper, Kurt A., Velcheti, Vamsidhar, Madabhushi, Anant. (2019). Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage non–small cell lung cancer. *Clinical Cancer Research*, *25*(5), 1526–1534.

Cross, Simon, Furness, Peter, Igali, Laszlo, Snead, David, Treanor, Darren. (2018). *Best practice recommendations for implementing digital pathology Unique document number G162 Document name Best practice recommendations for implementing digital pathology*.

Davenport, A., Hale, R. J., Hunt, C. R., Bigley, G., McMahon, R. F. T. (2003). Expression of Ki-67 and cytokeratin 20 in hyperplastic polyps of the colorectum. *Journal of Clinical Pathology*, *56*(3), 200–204.

De Jong, Koert P., Slellema, Rudi, Karrenbeld, Arend, Koudstaal, Jan, Gouw, Annette S. H., Sluiter, Wlm J., Peeters, Paul M. J. G., Slooff, Maarten J. H., De Vries, Elisabeth G. E. (1998). Clinical relevance of transforming growth factor α, epidermal growth factor receptor, p53, and Ki67 in colorectal liver metastases and corresponding primary tumors. *Hepatology*, *28*(4), 971–979.

de Mascarel, A., Coindre, J. M., de Mascarel, I., Trojani, M., Marée, D., Hoerni, B. (1981). The prognostic significance of specific histologic features of carcinoma of the colon and rectum. *Surgery, Gynecology & Obstetrics*, *153*(4), 511–514.

De Rosa, Nicole, Rodriguez-Bigas, Miguel A., Chang, George J., Veerapong, Jula, Borras, Ester, Krishnan, Sunil, Bednarski, Brian, Messick, Craig A., Skibber, John M., Feig, Barry W., Lynch, Patrick M., Vilar, Eduardo, You, Y. Nancy. (2016). DNA Mismatch Repair Deficiency in Rectal Cancer: Benchmarking Its Impact on Prognosis,

Neoadjuvant Response Prediction, and Clinical Cancer Genetics. *Journal of Clinical Oncology*, *34*(25), 3039.

De Smedt, L., Lemahieu, J., Palmans, S., Govaere, O., Tousseyn, T., Van Cutsem, E., Prenen, H., Tejpar, S., Spaepen, M., Matthijs, G., Decaestecker, C., Moles Lopez, X., Demetter, P., Salmon, I., Sagaert, X. (2015). Microsatellite instable vs stable colon carcinomas: analysis of tumour heterogeneity, inflammation and angiogenesis. *British Journal of Cancer 2015 113:3*, *113*(3), 500–509.

De Sousa E Melo, Felipe, Wang, Xin, Jansen, Marnix, Fessler, Evelyn, Trinh, Anne, De Rooij, Laura P. M. H., De Jong, Joan H., De Boer, Onno J., Van Leersum, Ronald, Bijlsma, Maarten F., Rodermond, Hans, Van Der Heijden, Maartje, Van Noesel, Carel J. M., Tuynman, Jurriaan B., Dekker, Evelien, Markowetz, Florian, Medema, Jan Paul, Vermeulen, Louis. (2013). Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nature Medicine 2013 19:5*, *19*(5), 614–618.

De Wever, Olivier, Van Bockstal, Mieke, Mareel, Marc, Hendrix, An, Bracke, Marc. (2014). Carcinoma-associated fibroblasts provide operational flexibility in metastasis. *Seminars in Cancer Biology*, *25*, 33–46.

Deng, Jia, Dong, Wei, Socher, Richard, Li, Li-Jia, Kai Li, Li Fei-Fei. (2010). *ImageNet: A large-scale hierarchical image database*. 248–255.

Denoix, PF. (1946). Enquete permanent dans les centres anticancereaux. *Bull Inst Natl Hyg*, *1*, 70–75.

Deschoolmeester, Vanessa, Baay, Marc, Lardon, Filip, Pauwel, Patrick, Peeters, Marc. (2011). Immune cells in colorectal cancer: Prognostic relevance and role of MSI. *Cancer Microenvironment*, *4*(3), 377–392.

Dice, Lee R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, *26*(3), 297–302.

Dienstmann, Rodrigo, Vermeulen, Louis, Guinney, Justin, Kopetz, Scott, Tejpar, Sabine, Tabernero, Josep. (2017). Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nature Reviews Cancer 2017 17:2*, *17*(2), 79–92.

Dukes, C. (1932). The classification of cancer of the rectum. *J Pathol Bacteriol*, *35*, 323–332.

Ellis, Harold, Mahadevan, Vishy. (2014). Anatomy of the caecum, appendix and colon. *Surgery (Oxford)*, *32*(4), 155–158.

Elomaa, Hanna, Ahtiainen, Maarit, Väyrynen, Sara A., Ogino, Shuji, Nowak, Jonathan A., Friman, Marjukka, Helminen, Olli, Wirta, Erkki Ville, Seppälä, Toni T., Böhm, Jan, Mäkinen, Markus J., Mecklin, Jukka Pekka, Kuopio, Teijo, Väyrynen, Juha P. (2022). Prognostic significance of spatial and density analysis of T lymphocytes in colorectal cancer. *British Journal of Cancer 2022 127:3*, *127*(3), 514–523.

Evans, Andrew J., Brown, Richard W., Bui, Marilyn M., Chlipala, Elizabeth A., Lacchetti, Christina, Milner, Danny A., Pantanowitz, Liron, Parwani, Anil V., Reid, Kearin, Riben, Michael W., Reuter, Victor E., Stephens, Lisa, Stewart, Rachel L., Thomas, Nicole E. (2022). Validating Whole Slide Imaging Systems for Diagnostic Purposes in PathologyGuideline Update From the College of American Pathologists in Collaboration With the American Society for Clinical Pathology and the Association for Pathology Informatics. *Archives of Pathology & Laboratory Medicine*, *146*(4), 440–450.

Falkenius, Johan, Johansson, Hemming, Tuominen, Rainer, Frostvik Stolt, Marianne, Hansson, Johan, Egyhazi Brage, Suzanne. (2017). Presence of immune cells, low tumor proliferation and wild type BRAF mutation status is associated with a favourable clinical outcome in stage III cutaneous melanoma. *BMC Cancer*, *17*(1), 1–9.

Fang, Minggang, Ou, Jianhong, Hutchinson, Lloyd, Green, Michael R. (2014). The BRAF Oncoprotein Functions through the Transcriptional Repressor MAFG to Mediate the CpG Island Methylator Phenotype. *Molecular Cell*, *55*(6), 904–915.

Fearon, Eric R., Vogelstein, Bert. (1990). A genetic model for colorectal tumorigenesis. *Cell*, *61*(5), 759–767.

Fernández-Medarde, Alberto, Santos, Eugenio. (2011). Ras in cancer and developmental diseases. *Genes and Cancer*, *2*(3), 344–358.

Flavell, Richard A., Sanjabi, Shomyseh, Wrzesinski, Stephen H., Licona-Limón, Paula. (2010). The polarization of immune cells in the tumour environment by TGFβ. *Nature Reviews Immunology 2010 10:8*, *10*(8), 554–567.

Fluge, O., Gravdal, K., Carlsen, E., Vonen, B., Kjellevold, K., Refsum, S., Lilleng, R., Eide, T. J., Halvorsen, T. B., Tveit, K. M., Otte, A. P., Akslen, L. A., Dahl, O. (2009).

Expression of EZH2 and Ki-67 in colorectal cancer and associations with treatment response and prognosis. *British Journal of Cancer 2009 101:8*, *101*(8), 1282–1289.

Freeman, Michael R., Li, Quanlin, Chung, Leland W. K. (2013). Can stroma reaction predict cancer lethality? *Clinical Cancer Research*, *19*(18), 4905–4907.

Fukushima, Kunihiko, Miyake, Sei. (1982). *Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition*. 267–285.

Gabbert, Helmut, Wagner, Rudolf, Moll, Roland, Gerharz, Claus Dieter. (1985). Tumor dedifferentiation: An important step in tumor invasion. *Clinical & Experimental Metastasis 1985 3:4*, *3*(4), 257–279.

Gabriel, W. B., Dukes, Cuthbert, Bussey, H. J. R. (1935). Lymphatic spread in cancer of the rectum. *British Journal of Surgery*, *23*(90), 395–413.

Galon, Jérôme, Costes, Anne, Sanchez-Cabo, Fatima, Kirilovsky, Amos, Mlecnik, Bernhard, Lagorce-Pagès, Christine, Tosolini, Marie, Camus, Matthieu, Berger, Anne, Wind, Philippe, Zinzindohoué, Franck, Bruneval, Patrick, Cugnenc, Paul Henri, Trajanoski, Zlatko, Fridman, Wolf Herman, Pagès, Franck. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*, *313*(5795), 1960–1964.

Galon, Jérôme, Mlecnik, Bernhard, Bindea, Gabriela, Angell, Helen K., Berger, Anne, Lagorce, Christine, Lugli, Alessandro, Zlobec, Inti, Hartmann, Arndt, Bifulco, Carlo, Nagtegaal, Iris D., Palmqvist, Richard, Masucci, Giuseppe V., Botti, Gerardo, Tatangelo, Fabiana, Delrio, Paolo, Maio, Michele, Laghi, Luigi, Grizzi, Fabio, Asslaber, Martin, D'Arrigo, Corrado, Vidal-Vanaclocha, Fernando, Zavadova, Eva, Chouchane, Lotfi, Ohashi, Pamela S., Hafezi-Bakhtiari, Sara, Wouters, Bradly G., Roehrl, Michael, Nguyen, Linh, Kawakami, Yutaka, Hazama, Shoichi, Okuno, Kiyotaka, Ogino, Shuji, Gibbs, Peter, Waring, Paul, Sato, Noriyuki, Torigoe, Toshihiko, Itoh, Kyogo, Patel, Prabhu S., Shukla, Shilin N., Wang, Yili, Kopetz, Scott, Sinicrope, Frank A., Scripcariu, Viorel, Ascierto, Paolo A., Marincola, Francesco M., Fox, Bernard A., Pagès, Franck. (2014). Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. In *Journal of Pathology* (Vol. 232, Issue 2, pp. 199–209). John Wiley & Sons, Ltd.

Gao, Xian Hua, Yu, Guan Yu, Gong, Hai Feng, Liu, Lian Jie, Xu, Yi, Hao, Li Qiang, Liu, Peng, Liu, Zhi Hong, Bai, Chen Guang, Zhang, Wei. (2017). Differences of protein

expression profiles, KRAS and BRAF mutation, and prognosis in right-sided colon, left-sided colon and rectal cancer. *Scientific Reports*, *7*(1).

García-Solano, José, Pérez-Guillermo, Miguel, Conesa-Zamora, Pablo, Acosta-Ortega, Jesús, Trujillo-Santos, Javier, Cerezuela-Fuentes, Pablo, Mäkinen, Markus J. (2010). Clinicopathologic study of 85 colorectal serrated adenocarcinomas: further insights into the full recognition of a new subset of colorectal carcinoma. *Human Pathology*, *41*(10), 1359–1368.

Geessink, Oscar G. F., Baidoshvili, Alexi, Klaase, Joost M., Ehteshami Bejnordi, Babak, Litjens, Geert J. S., van Pelt, Gabi W., Mesker, Wilma E., Nagtegaal, Iris D., Ciompi, Francesco, van der Laak, Jeroen A. W. M. (2019). Computer aided quantification of intratumoral stroma yields an independent prognosticator in rectal cancer. *Cellular Oncology*, *42*(3), 331–341.

Genç, C. G., Falconi, M., Partelli, S., Muffatti, F., van Eeden, S., Doglioni, C., Klümpen, H. J., van Eijck, C. H. J., Nieveen van Dijkum, E. J. M. (2018). Recurrence of Pancreatic Neuroendocrine Tumors and Survival Predicted by Ki67. *Annals of Surgical Oncology*, *25*(8), 2467–2474.

Gerdes, J., Lemke, H., Baisch, H., Wacker, H. H., Schwab, U., Stein, H. (1984). Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. *The Journal of Immunology*, *133*(4).

Ginty, Fiona, Adak, Sudeshna, Can, Ali, Gerdes, Michael, Larsen, Melinda, Cline, Harvey, Filkins, Robert, Pang, Zhengyu, Li, Qing, Montalto, Michael C. (2008). The Relative Distribution of Membranous and Cytoplasmic Met Is a Prognostic Indicator in Stage I and II Colon Cancer. *Clinical Cancer Research*, *14*(12), 3814–3822.

Goetz, Jacky G., Minguet, Susana, Navarro-Lérida, Inmaculada, Lazcano, Juan José, Samaniego, Rafael, Calvo, Enrique, Tello, Marta, Osteso-Ibáñez, Teresa, Pellinen, Teijo, Echarri, Asier, Cerezo, Ana, Klein-Szanto, Andres J. P., Garcia, Ricardo, Keely, Patricia J., Sánchez-Mateos, Paloma, Cukierman, Edna, Del Pozo, Miguel A. (2011). Biomechanical Remodeling of the Microenvironment by Stromal Caveolin-1 Favors Tumor Invasion and Metastasis. *Cell*, *146*(1), 148–163.

Goldhirsch, A., Winer, E. P., Coates, A. S., Gelber, R. D., Piccart-Gebhart, M., Thürlimann, B., Senn, H. J., Albain, Kathy S., André, Fabrice, Bergh, Jonas, Bonnefoi, Hervé, Bretel-Morales, Denisse, Burstein, Harold, Cardoso, Fatima, Castiglione-Gertsch,

Monica, Coates, Alan S., Colleoni, Marco, Costa, Alberto, Curigliano, Giuseppe, Davidson, Nancy E., Leo, Angelo Di, Ejlertsen, Bent, … Wood, William C. (2013). Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Annals of Oncology*, *24*(9), 2206–2223.

Goldhirsch, A., Wood, W. C., Coates, A. S., Gelber, R. D., Thürlimann, B., Senn, H. J. (2011). Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of Oncology*, *22*(8), 1736–1747.

Graham, Simon, Chen, Hao, Gamper, Jevgenij, Dou, Qi, Heng, Pheng Ann, Snead, David, Tsang, Yee Wah, Rajpoot, Nasir. (2019). MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Medical Image Analysis*, *52*, 199–211.

Greenson, Joel K., Bonner, Joseph D., Ben-Yzhak, Ofer, Cohen, Hector I., Miselevich, Ines, Resnick, Murray B., Trougouboff, Philippe, Tomsho, Lynn D., Kim, Evelyn, Low, Marcelo, Almog, Ronit, Rennert, Gad, Gruber, Stephen B. (2003). Phenotype of microsatellite unstable colorectal carcinomas: Well-differentiated and focally mucinous tumors and the absence of dirty necrosis correlate with microsatellite instability. *The American Journal of Surgical Pathology*, *27*(5), 563–570.

Guilherme Campos, Fabio. (2014). Surgical treatment of familial adenomatous polyposis: Dilemmas and current recommendations. *World Journal of Gastroenterology : WJG*, *20*(44), 16620.

Guinney, Justin, Dienstmann, Rodrigo, Wang, Xin, de Reyni, lien, Schlicker, Andreas, Soneson, Charlotte, Marisa, Laetitia, Roepman, Paul, Nyamundanda, Gift, Angelino, Paolo, Bot, Brian M., Morris, Jeffrey S., Simon, Iris M., Gerster, Sarah, Fessler, Evelyn, De Sousa Melo, Felipe E., Missiaglia, Edoardo, Ramay, Hena, Barras, David, Homicsko, Krisztian, Maru, Dipen, Manyam, Ganiraju C., Broom, Bradley, Boige, Valerie, Perez-Villamil, Beatriz, Laderas, Ted, Salazar, Ramon, Gray, Joe W., Hanahan, Douglas, Tabernero, Josep, Bernards, Rene, Friend, Stephen H., Laurent-Puig, Pierre, Paul Medema, Jan, Sadanandam, Anguraj, Wessels, Lodewyk, Delorenzi, Mauro, Kopetz, Scott, Guinney, Justin, Vermeulen, Louis, Tejpar, Sabine. (2015). The consensus molecular subtypes of colorectal cancer. *Nature Medicine*.

Gujam, F. J. A., McMillan, D. C., Mohammed, Z. M. A., Edwards, J., Going, J. J. (2015). The relationship between tumour budding, the tumour microenvironment and survival in patients with invasive ductal breast cancer. *British Journal of Cancer 2015 113:7*, *113*(7), 1066–1074.

Gurcan, Metin N., Boucheron, Laura E., Can, Ali, Madabhushi, Anant, Rajpoot, Nasir M., Yener, Bulent. (2009). Histopathological Image Analysis: A Review. *IEEE Reviews in Biomedical Engineering*, *2*, 147–171.

Hacking, Sean M., Wu, Dongling, Alexis, Claudine, Nasim, Mansoor. (2022). A Novel Superpixel Approach to the Tumoral Microenvironment in Colorectal Cancer. *Journal of Pathology Informatics*, *13*, 100009.

Halvarsson, Britta, Anderson, Harald, Domanska, Katarina, Lindmark, Gudrun, Nilhert, Mef. (2008). Clinicopathologic factors identify sporadic mismatch repair-defective colon cancers. *American Journal of Clinical Pathology*, *129*(2), 238–244.

Halvorsen, T. B., Seim, E. (1988). Degree of differentiation in colorectal adenocarcinomas: a multivariate analysis of the influence on survival. *Journal of Clinical Pathology*, *41*(5), 532.

Halvorsen, T. B., Seimt, Eva. (1989). Association between invasiveness, inflammatory reaction, desmoplasia and survival in colorectal cancer. *J Clin Pathol*, *42*, 162–166.

Hanahan, Douglas, Weinberg, Robert A. (2011). Hallmarks of cancer: The next generation. *Cell*, *144*(5), 646–674.

Hanna, Matthew G., Reuter, Victor E., Hameed, Meera R., Tan, Lee K., Chiang, Sarah, Sigel, Carlie, Hollmann, Travis, Giri, Dilip, Samboy, Jennifer, Moradel, Carlos, Rosado, Andrea, Otilano, John R., England, Christine, Corsale, Lorraine, Stamelos, Evangelos, Yagi, Yukako, Schüffler, Peter J., Fuchs, Thomas, Klimstra, David S., Sirintrapun, S. Joseph. (2019). Whole slide imaging equivalency and efficiency study: experience at a large academic center. *Modern Pathology*, *32*(7), 916–928.

Hasebe, Takahiro, Sasaki, Satoshi, Imoto, Shigeru, Ochiai, Atsushi. (2001). Highly Proliferative Fibroblasts Forming Fibrotic Focus Govern Metastasis of Invasive Ductal Carcinoma of the Breast. *Modern Pathology 2001 14:4*, *14*(4), 325–337.

Hashimoto, Mai, Uesugi, Noriyuki, Sugai, Mayu, Ito, Kazuhiro, Yanagawa, Naoki, Otsuka, Koki, Kajiwara, Yoshiki, Ueno, Hideki, Sasaki, Akira, Sugai, Tamotsu. (2022). Desmoplastic reactions and epithelial-mesenchymal transition proteins in stages II and

III colorectal cancer: association with and prognostic value for disease-free survival. *Virchows Archiv*, *480*(4), 793–805.

He, Tong Chuan, Sparks, Andrew B., Rago, Carlo, Hermeking, Heiko, Zawel, Leigh, Da Costa, Luis T., Morin, Patrice J., Vogelstein, Bert, Kinzler, Kenneth W. (1998). Identification of c-MYC as a target of the APC pathway. *Science (New York, N.Y.)*, *281*(5382), 1509–1512.

Helder-Woolderink, J. M., Blok, E. A., Vasen, H. F. A., Hollema, H., Mourits, M. J., De Bock, G. H. (2016). Ovarian cancer in Lynch syndrome; a systematic review. *European Journal of Cancer*, *55*, 65–73.

Heneberg, Petr. (2016). Paracrine tumor signaling induces transdifferentiation of surrounding fibroblasts. *Critical Reviews in Oncology/Hematology*, *97*, 303–311.

Hinton, Geoffrey E., Srivastava, Nitish, Krizhevsky, Alex, Sutskever, Ilya, Salakhutdinov, Ruslan R. (2012). *Improving neural networks by preventing co-adaptation of feature detectors*.

Hinz, Boris, Phan, Sem H., Thannickal, Victor J., Prunotto, Marco, Desmoulire, Alexis, Varga, John, De Wever, Olivier, Mareel, Marc, Gabbiani, Giulio. (2012). Recent Developments in Myofibroblast Biology: Paradigms for Connective Tissue Remodeling. *The American Journal of Pathology*, *180*(4), 1340–1355.

Ho, David Joon, Yarlagadda, Dig V. K., D'Alfonso, Timothy M., Hanna, Matthew G., Grabenstetter, Anne, Ntiamoah, Peter, Brogi, Edi, Tan, Lee K., Fuchs, Thomas J. (2021). Deep Multi-Magnification Networks for multi-class breast cancer image segmentation. *Computerized Medical Imaging and Graphics*, *88*, 101866.

Hörkkö, T. T., Mäkinen, M. J. (2003). Colorectal proliferation and apoptosis in serrated versus conventional adenoma-carcinoma pathway: growth, progression and survival. *Scandinavian Journal of Gastroenterology*, *38*(12), 1241–1248.

Howard, Andrew G., Zhu, Menglong, Chen, Bo, Kalenichenko, Dmitry, Wang, Weijun, Weyand, Tobias, Andreetto, Marco, Adam, Hartwig. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*.

Howlett, C. J., Tweedie, E. J., Driman, D. K. (2009). Use of an elastic stain to show venous invasion in colorectal carcinoma: a simple technique for detection of an important prognostic factor. *Journal of Clinical Pathology*, *62*(11), 1021–1025.

Hugen, N., Verhoeven, R. H. A., Radema, S. A., de Hingh, I. H. J. T., Pruijt, J. F. M., Nagtegaal, I. D., Lemmens, V. E. P. P., de Wilt, J. H. W. (2013). Prognosis and value of adjuvant chemotherapy in stage III mucinous colorectal carcinoma. *Annals of Oncology*, *24*(11), 2819–2824.

Hugen, Niek, Brown, Gina, Glynne-Jones, Robert, De Wilt, Johannes H. W., Nagtegaal, Iris D. (2015). Advances in the care of patients with mucinous colorectal cancer. *Nature Reviews Clinical Oncology 2015 13:6*, *13*(6), 361–369.

Huh, Jung Wook, Kim, Hyeong Rok, Kim, Young Jin. (2010). Lymphovascular or perineural invasion may predict lymph node metastasis in patients with T1 and T2 colorectal cancer. *Journal of Gastrointestinal Surgery*, *14*(7), 1074–1080.

Huh, Jung Wook, Lee, Jae Hyuk, Kim, Hyeong Rok. (2012). Prognostic Significance of Tumor-Infiltrating Lymphocytes for Patients With Colorectal Cancer. *Archives of Surgery*, *147*(4), 366–372.

Huijbers, A. (2013). The proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients: validation in the VICTOR trial. *Ann Oncol*, *24*(1), 179–185.

Huijbers, A., Tollenaar, R. A. E. M., Pelt, G. W. V., Zeestraten, E. C. M., Dutton, S., McConkey, C. C., Domingo, E., Smit, V. T. H. B. M., Midgley, R., Warren, B. F., Johnstone, E. C., Kerr, D. J., Mesker, W. E. (2013). The proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients: Validation in the victor trial. *Annals of Oncology*, *24*(1), 179–185.

Hynes, Seán O., Coleman, Helen G., Kelly, Paul J., Irwin, Steven, O'Neill, Roisin F., Gray, Ronan T., McGready, Claire, Dunne, Philip D., McQuaid, Stephen, James, Jacqueline A., Salto-Tellez, Manuel, Loughrey, Maurice B. (2017a). Back to the future: routine morphological assessment of the tumour microenvironment is prognostic in stage II/III colon cancer in a large population-based study. *Histopathology*, *71*(1), 12–26.

Hynes, Seán O., Coleman, Helen G., Kelly, Paul J., Irwin, Steven, O'Neill, Roisin F., Gray, Ronan T., McGready, Claire, Dunne, Philip D., McQuaid, Stephen, James, Jacqueline A., Salto-Tellez, Manuel, Loughrey, Maurice B. (2017b). Back to the future: routine morphological assessment of the tumour microenvironment is prognostic in stage II/III colon cancer in a large population-based study. *Histopathology*, *71*(1), 12–26.

Hynes, Seán O., Coleman, Helen G., Kelly, Paul J., Irwin, Steven, O'Neill, Roisin F., Gray, Ronan T., McGready, Claire, Dunne, Philip D., McQuaid, Stephen, James, Jacqueline A., Salto-Tellez, Manuel, Loughrey, Maurice B. (2017c). Back to the future: routine morphological assessment of the tumour microenvironment is prognostic in stage II/III colon cancer in a large population-based study. *Histopathology*, *71*(1), 12–26.

Imai, T. (1954). The growth of human carcinoma: a morphological analysis. *Fukuoka Igaku Zasshi*, *45*, 30.

Ionov, Y., Peinado, MA, Malkhosyan, S., Nature, D. Shibata-, 1993, undefined. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature.Com*.

Isella, Claudio, Brundu, Francesco, Bellomo, Sara E., Galimi, Francesco, Zanella, Eugenia, Porporato, Roberta, Petti, Consalvo, Fiori, Alessandro, Orzan, Francesca, Senetta, Rebecca, Boccaccio, Carla, Ficarra, Elisa, Marchionni, Luigi, Trusolino, Livio, Medico, Enzo, Bertotti, Andrea. (2017a). Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. *Nature Communications 2017 8:1*, *8*(1), 1–16.

Isella, Claudio, Brundu, Francesco, Bellomo, Sara E., Galimi, Francesco, Zanella, Eugenia, Porporato, Roberta, Petti, Consalvo, Fiori, Alessandro, Orzan, Francesca, Senetta, Rebecca, Boccaccio, Carla, Ficarra, Elisa, Marchionni, Luigi, Trusolino, Livio, Medico, Enzo, Bertotti, Andrea. (2017b). Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. *Nature Communications 2017 8:1*, *8*(1), 1–16.

Isella, Claudio, Terrasi, Andrea, Bellomo, Sara Erika, Petti, Consalvo, Galatola, Giovanni, Muratore, Andrea, Mellano, Alfredo, Senetta, Rebecca, Cassenti, Adele, Sonetto, Cristina, Inghirami, Giorgio, Trusolino, Livio, Fekete, Zsolt, De Ridder, Mark, Cassoni, Paola, Storme, Guy, Bertotti, Andrea, Medico, Enzo. (2015). Stromal contribution to the colorectal cancer transcriptome. *Nature Genetics 2015 47:4*, *47*(4), 312–319.

Issa, Jean Pierre. (2004). CpG island methylator phenotype in cancer. *Nature Reviews Cancer 2004 4:12*, *4*(12), 988–993.

Jass, J. R. (1986). Lymphocytic infiltration and survival in rectal cancer. *Journal of Clinical Pathology*, *39*(6), 585–589.

Jass, J. R. (2007). Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology*, *50*(1), 113–130.

Jass, J. R., Love, S. B., Northover, J. M. A. (1987). A NEW PROGNOSTIC CLASSIFICATION OF RECTAL CANCER. *The Lancet*, *329*(8545), 1303–1306.

Jass, J. R., Smith, M. (1992). Sialic acid and epithelial differentiation in colorectal polyps and cancer — a morphological, mucin and lectin histochemical study. *Pathology*, *24*(4), 233–242.

Jia, Min, Gao, Xu, Zhang, Yan, Hoffmeister, Michael, Brenner, Hermann. (2016). Different definitions of CpG island methylator phenotype and outcomes of colorectal cancer: a systematic review. *Clinical Epigenetics*, *8*(1), 1–14.

Jung, Andreas, Schrauder, Michael, Oswald, Ursula, Knoll, Claudia, Sellberg, Petter, Palmqvist, Richard, Niedobitek, Gerald, Brabletz, Thomas, Kirchner, Thomas. (2001). The Invasion Front of Human Colorectal Adenocarcinomas Shows Co-Localization of Nuclear β-Catenin, Cyclin D1, and p16INK4A and Is a Region of Low Proliferation. *The American Journal of Pathology*, *159*(5), 1613–1617.

Jung, Minsun, Lee, Ji Ae, Yoo, Seung Yeon, Bae, Jeong Mo, Kang, Gyeong Hoon, Kim, Jung Ho. (2022). Intratumoral spatial heterogeneity of tumor-infiltrating lymphocytes is a significant factor for precisely stratifying prognostic immune subgroups of microsatellite instability-high colorectal carcinomas. *Modern Pathology 2022*, 1–12.

Kane, Michael F., Loda, Massimo, Gaida, Gretchen M., Lipman, Jennifer, Mishra, Rajesh, Goldman, Harvey, Milburnjessup, J., Dana, Charles A. (1997). Methylation of the hMLH1 Promoter Correlates with Lack of Expression of hMLH1 in Sporadic Colon Tumors and Mismatch Repair-defective Human Tumor Cell Lines. *AACR*, *57*, 808–881.

Kaneda, Atsushi, Yagi, Koichi. (2011). Two groups of DNA methylation markers to classify colorectal cancer into three epigenotypes. *Cancer Science*, *102*(1), 18–24.

Kang, Hakjung, O'Connell, Jessica B., Maggard, Melinda A., Sack, Jonathan, Ko, Clifford Y. (2005). A 10-year outcomes evaluation of mucinous and signet-ring cell carcinoma of the colon and rectum. *Diseases of the Colon and Rectum*, *48*(6), 1161–1168.

Kather, Jakob Nikolas, Krisam, Johannes, Charoentong, Pornpimol, Luedde, Tom, Herpel, Esther, Weis, Cleo Aron, Gaiser, Timo, Marx, Alexander, Valous, Nektarios A., Ferber, Dyke, Jansen, Lina, Reyes-Aldasoro, Constantino Carlos, Zörnig, Inka, Jäger, Dirk, Brenner, Hermann, Chang-Claude, Jenny, Hoffmeister, Michael, Halama, Niels. (2019).

Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLOS Medicine*, *16*(1), e1002730.

Kather, Jakob Nikolas, Pearson, Alexander T., Halama, Niels, Jäger, Dirk, Krause, Jeremias, Loosen, Sven H., Marx, Alexander, Boor, Peter, Tacke, Frank, Neumann, Ulf Peter, Grabsch, Heike I., Yoshikawa, Takaki, Brenner, Hermann, Chang-Claude, Jenny, Hoffmeister, Michael, Trautwein, Christian, Luedde, Tom. (2019). Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nature Medicine*, *25*(7), 1054.

Kather, Jakob Nikolas, Weis, Cleo Aron, Bianconi, Francesco, Melchers, Susanne M., Schad, Lothar R., Gaiser, Timo, Marx, Alexander, Zöllner, Frank Gerrit. (2016). Multi-class texture analysis in colorectal cancer histology. *Scientific Reports 2016 6:1*, *6*(1), 1–11.

Kingma, Diederik P., Lei Ba, Jimmy. (n.d.). *ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION*.

Kinzler, Kenneth W., Vogelstein, Bert. (1996). Lessons from hereditary colorectal cancer. *Cell*, *87*(2), 159–170.

Klintrup, Kai, Mäkinen, Johanna M., Kauppila, Saila, Väre, Päivi O., Melkko, Jukka, Tuominen, Hannu, Tuppurainen, Karoliina, Mäkelä, Jyrki, Karttunen, Tuomo J., Mäkinen, Markus J. (2005). Inflammation and prognosis in colorectal cancer. *European Journal of Cancer*, *41*(17), 2645–2654.

Klöppel, Günter, La Rosa, Stefano. (2018). Ki67 labeling index: assessment and prognostic role in gastroenteropancreatic neuroendocrine neoplasms. *Virchows Archiv*, *472*(3), 341–349.

Kocián, Petr, Šedivcová, Monika, Drgáč, Jan, Černá, Kateřina, Hoch, Jiří, Kodet, Roman, Bartůňková, Jiřina, Špíšek, Radek, Fialová, Anna. (2011). Tumor-infiltrating lymphocytes and dendritic cells in human colorectal cancer: Their relationship to KRAS mutational status and disease recurrence. *Human Immunology*, *72*(11), 1022–1028.

Komura, Daisuke, Ishikawa, Shumpei. (2018). Machine Learning Methods for Histopathological Image Analysis. *Computational and Structural Biotechnology Journal*, *16*, 34–42.

Koopman, Timco, de Bock, Geertruida H., Buikema, Henk J., Smits, Maria M., Louwen, Maarten, Hage, Mariska, Imholz, Alex L. T., van der Vegt, Bert. (2018). Digital image

analysis of HER2 immunohistochemistry in gastric- and oesophageal adenocarcinoma: a validation study on biopsies and surgical specimens. *Histopathology*, *72*(2), 191–200.

Koornstra, Jan J., Mourits, Marian JE, Sijmons, Rolf H., Leliveld, Annemarie M., Hollema, Harry, Kleibeuker, Jan H. (2009). Management of extracolonic tumours in patients with Lynch syndrome. *The Lancet Oncology*, *10*(4), 400–408.

Kramer, C. J. H., Vangangelt, K. M. H., van Pelt, G. W., Dekker, T. J. A., Tollenaar, R. A. E. M., Mesker, W. E. (2019). The prognostic value of tumour–stroma ratio in primary breast cancer with special attention to triple-negative tumours: a review. *Breast Cancer Research and Treatment*, *173*(1), 55–64.

Krizhevsky, Alex, Sutskever, Ilya, Hinton, Geoffrey E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, *25*.

Langner, Cord, Harbaum, Lars, Pollheimer, Marion J., Kornprat, Peter, Lindtner, Richard A., Schlemmer, Andrea, Vieth, Michael, Rehak, Peter. (2012). Mucinous differentiation in colorectal cancer – indicator of poor prognosis? *Histopathology*, *60*(7), 1060–1072.

Lara, Haydee, Li, Zaibo, Abels, Esther, Aeffner, Famke, Bui, Marilyn M., Elgabry, Ehab A., Kozlowski, Cleopatra, Montalto, Michael C., Parwani, Anil V., Zarella, Mark D., Bowman, Douglas, Rimm, David, Pantanowitz, Liron. (2021). Quantitative Image Analysis for Tissue Biomarker Use: A White Paper From the Digital Pathology Association. *Applied Immunohistochemistry & Molecular Morphology*, *29*(7), 479.

Laufman, H., Saphir, O. (1951). PRIMARY LINITIS PLASTICA TYPE OF CARCINOMA OF THE COLON. *A.M.A. Archives of Surgery*, *62*(1), 79–91.

LeCun, Yann, Bottou, Léon, Bengio, Yoshua, Haffner, Patrick. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Lenz, Heinz Josef, Ou, Fang Shu, Venook, Alan P., Hochster, Howard S., Niedzwiecki, Donna, Goldberg, Richard M., Mayer, Robert J., Bertagnolli, Monica M., Blanke, Charles D., Zemla, Tyler, Qu, Xueping, Wirapati, Pratyaksha, Tejpar, Sabine, Innocenti, Federico, Kabbarah, Omar. (2019). Impact of Consensus Molecular Subtype on Survival in Patients With Metastatic Colorectal Cancer: Results From CALGB/SWOG 80405 (Alliance). *Journal of Clinical Oncology*, *37*(22), 1876.

Leppänen, Joni, Lindholm, Ville, Isohookana, Joel, Haapasaari, Kirsi Maria, Karihtala, Peeter, Lehenkari, Petri P., Saarnio, Juha, Kauppila, Joonas H., Karttunen, Tuomo J., Helminen, Olli, Huhta, Heikki. (2019). Tenascin C, Fibronectin, and Tumor-Stroma Ratio in Pancreatic Ductal Adenocarcinoma. *Pancreas*, *48*(1), 43.

Leslie, A., Carey, F. A., Pratt, N. R., Steele, R. J. C. (2002). The colorectal adenoma-carcinoma sequence. *British Journal of Surgery*, *89*(7), 845–860.

Li, H., Yuan, S. L., Han, Z. Z., Huang, J., Cui, L., Jiang, C. Q., Zhang, Y. (2017). Prognostic significance of the tumor-stroma ratio in gallbladder cancer. *588 Neoplasma*, *64*.

Li, Zaibo, Bui, Marilyn M., Pantanowitz, Liron. (2022). Clinical tissue biomarker digital image analysis: A review of current applications. *Human Pathology Reports*, *28*, 300633.

Liu, Jiao, Kang, Rui, Tang, Daolin. (2021). The KRAS-G12C inhibitor: activity and resistance. *Cancer Gene Therapy 2021 29:7*, *29*(7), 875–878.

Liu, Jing, Liu, Juan, Li, Jinsong, Chen, Yingling, Guan, Xiaoling, Wu, Xiaojuan, Hao, Chunyan, Sun, Yanlin, Wang, Yan, Wang, Xiao. (2014). Tumor–stroma ratio is an independent predictor for survival in early cervical carcinoma. *Gynecologic Oncology*, *132*(1), 81–86.

Loughrey, Maurice B., Webster, Fleur, Arends, Mark J., Brown, Ian, Burgart, Lawrence J., Cunningham, Chris, Flejou, Jean Francois, Kakar, Sanjay, Kirsch, Richard, Kojima, Motohiro, Lugli, Alessandro, Rosty, Christophe, Sheahan, Kieran, West, Nicholas P., Wilson, Richard H., Nagtegaal, Iris D. (2022). Dataset for Pathology Reporting of Colorectal Cancer: Recommendations From the International Collaboration on Cancer Reporting (ICCR). *Annals of Surgery*, *275*(3), e549.

Lu, Zixiao, Xu, Siwen, Shao, Wei, Wu, Yi, Zhang, Jie, Han, Zhi, Feng, Qianjin, Huang, Kun. (2020). Deep-Learning–Based Characterization of Tumor-Infiltrating Lymphocytes in Breast Cancers From Histopathology Images and Multiomics Data. *JCO Clinical Cancer Informatics*, *4*, 480–490.

Lugli, A., Zlobec, I., Minoo, P., Baker, K., Tornillo, L., Terracciano, L., Jass, J. R. (2007). Prognostic significance of the wnt signalling pathway molecules APC, β-catenin and E-cadherin in colorectal cancer—a tissue microarray-based analysis. *Histopathology*, *50*(4), 453–464.

Lugli, Alessandro, Kirsch, Richard, Ajioka, Yoichi, Bosman, Fred, Cathomas, Gieri, Dawson, Heather, El Zimaity, Hala, Fléjou, Jean François, Hansen, Tine Plato, Hartmann, Arndt, Kakar, Sanjay, Langner, Cord, Nagtegaal, Iris, Puppa, Giacomo, Riddell, Robert, Ristimäki, Ari, Sheahan, Kieran, Smyrk, Thomas, Sugihara, Kenichi, Terris, Benoît, Ueno, Hideki, Vieth, Michael, Zlobec, Inti, Quirke, Phil. (2017). Recommendations for reporting tumor budding in colorectal cancer based on the International Tumor Budding Consensus Conference (ITBCC) 2016. *Modern Pathology 2017 30:9*, *30*(9), 1299–1311.

Lugli, Alessandro, Kirsch, Richard, Ajioka, Yoichi, Bosman, Fred, Cathomas, Gieri, Dawson, Heather, Zimaity, Hala El, Fléjou, Jean François, Hansen, Tine Plato, Hartmann, Arndt, Kakar, Sanjay, Langner, Cord, Nagtegaal, Iris, Puppa, Giacomo, Riddell, Robert, Ristimäki, Ari, Sheahan, Kieran, Smyrk, Thomas, Sugihara, Kenichi, Terris, Benoît, Ueno, Hideki, Vieth, Michael, Zlobec, Inti, Quirke, Phil. (2017). Recommendations for reporting tumor budding in colorectal cancer based on the International Tumor Budding Consensus Conference (ITBCC) 2016. *Modern Pathology 2017 30:9*, *30*(9), 1299–1311.

Luo, Zhao Wen, Zhu, Ming Gu, Zhang, Zhi Qiao, Ye, Feng Jun, Huang, Wen Heng, Luo, Xue Zhang. (2019). Increased expression of Ki-67 is a poor prognostic marker for colorectal cancer patients: A meta analysis. *BMC Cancer*, *19*(1), 1–13.

Lv, Zhen, Cai, Xianlei, Weng, Xiaoyu, Xiao, Heng, Du, Chengli, Cheng, Jun, Zhou, Lin, Xie, Haiyang, Sun, Ke, Wu, Jian, Zheng, Shusen. (2015). Tumor–stroma ratio is a prognostic factor for survival in hepatocellular carcinoma patients after liver resection or transplantation. *Surgery*, *158*(1), 142–150.

Ma, Handong, Wang, Jing, Zhao, Xueli, Wu, Tiantian, Huang, Zhengjie, Chen, Dafan, Liu, Yingfu, Ouyang, Gaoliang. (2020). Periostin Promotes Colorectal Tumorigenesis through Integrin-FAK-Src Pathway-Mediated YAP/TAZ Activation. *Cell Reports*, *30*(3), 793-806.e6.

Maccarty, William Carpenter. (1931a). PRINCIPLES OF PROGNOSIS IN CANCER. *Journal of the American Medical Association*, *96*(1), 30–33.

Maccarty, William Carpenter. (1931b). PRINCIPLES OF PROGNOSIS IN CANCER. *JAMA*, *96*(1), 30–33.

Macenko, Marc, Niethammer, Marc, Marron, J. S., Borland, David, Woosley, John T., Guan, Xiaojun, Schmitt, Charles, Thomas, Nancy E. (2009). A method for normalizing histology slides for quantitative analysis. *Proceedings - 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009*, 1107–1110.

Mäkinen, M. J. (2007). Colorectal serrated adenocarcinoma. *Histopathology*, *50*(1), 131–150.

Mansouri, D., McMillan, D. C., McIlveen, E., Crighton, E. M., Morrison, D. S., Horgan, P. G. (2016). A comparison of tumour and host prognostic factors in screen-detected vs nonscreen-detected colorectal cancer: a contemporaneous study. *Colorectal Disease*, *18*(10), 967–975.

Marisa, Laetitia, de Reyniès, Aurélien, Duval, Alex, Selves, Janick, Gaub, Marie Pierre, Vescovo, Laure, Etienne-Grimaldi, Marie Christine, Schiappa, Renaud, Guenot, Dominique, Ayadi, Mira, Kirzin, Sylvain, Chazal, Maurice, Fléjou, Jean François, Benchimol, Daniel, Berger, Anne, Lagarde, Arnaud, Pencreach, Erwan, Piard, Françoise, Elias, Dominique, Parc, Yann, Olschwang, Sylviane, Milano, Gérard, Laurent-Puig, Pierre, Boige, Valérie. (2013). Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value. *PLOS Medicine*, *10*(5), e1001453.

Marks, K. M., West, N. P., Morris, E., Quirke, P. (2018). Clinicopathological, genomic and immunological factors in colorectal cancer prognosis. *British Journal of Surgery*, *105*(2), e99–e109.

Martin, Benedikt, Banner, Bettina Monika, Schäfer, Eva Maria, Mayr, Patrick, Anthuber, Matthias, Schenkirsch, Gerhard, Märkl, Bruno. (2020). Tumor proportion in colon cancer: results from a semiautomatic image analysis approach. *Virchows Archiv*, *477*(2), 185.

McCulloch, Warren S., Pitts, Walter. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics 1943 5:4*, *5*(4), 115–133.

Melling, N., Kowitz, C. M., Simon, R. (2016). High Ki67 expression is an independent good prognostic marker in colorectal cancer. *J Clin Pathol*, *69*, 209–214.

Melling, Nathaniel, Kowitz, Charlotte Marie, Simon, Ronald, Bokemeyer, Carsten, Terracciano, Luigi, Sauter, Guido, Izbicki, Jakob Robert, Marx, Andreas Holger.

(2016). High Ki67 expression is an independent good prognostic marker in colorectal cancer. *Journal of Clinical Pathology*, *69*(3), 209–214.

Merlos-Suárez, Anna, Barriga, Francisco M., Jung, Peter, Iglesias, Mar, Céspedes, María Virtudes, Rossell, David, Sevillano, Marta, Hernando-Momblona, Xavier, Da Silva-Diz, Victoria, Muñoz, Purificación, Clevers, Hans, Sancho, Elena, Mangues, Ramón, Batlle, Eduard. (2011). The Intestinal Stem Cell Signature Identifies Colorectal Cancer Stem Cells and Predicts Disease Relapse. *Cell Stem Cell*, *8*(5), 511–524.

Mesker, Wilma E., Junggeburt, Jan M. C., Szuhai, Karoly, De Heer, Pieter, Morreau, Hans, Tanke, Hans J., Tollenaar, Rob A. E. M. (2007). The carcinoma-stromal ratio of colon carcinoma is an independent factor for survival compared to lymph node status and tumor stage. In *Cellular Oncology* (Vol. 29). IOS Press.

Mesker, Wilma E., Liefers, Gerrit Jan, Junggeburt, Jan M. C., Van Pelt, Gabi W., Alberici, Paola, Kuppen, Peter J. K., Miranda, Noel F., Van Leeuwen, Karin A. M., Morreau, Hans, Szuhai, Karoly, Tollenaar, Rob A. E. M., Tanke, Hans J. (2009). Presence of a high amount of stroma and downregulation of SMAD4 predict for worse survival for stage I-II colon cancer patients. *Cellular Oncology : The Official Journal of the International Society for Cellular Oncology*, *31*(3), 169–178.

Meskó, Bertalan, Görög, Marton. (2020). A short guide for medical professionals in the era of artificial intelligence. *Npj Digital Medicine 2020 3:1*, *3*(1), 1–8.

Messenger, David E., Driman, David K., Kirsch, Richard. (2012). Developments in the assessment of venous invasion in colorectal cancer: implications for future practice and patient outcome. *Human Pathology*, *43*(7), 965–973.

Michael-Robinson, Julie M., Reid, Lynne E., Purdie, David M., Biemer-Hü Ttmann, A. E., Walsh, Michael D., Pandeya, Nirmala, Simms, Lisa A., Young, Joanne P., Leggett, Barbara A., Jass, Jeremy R., Radford-Smith, Graham L. (2001). Proliferation, Apoptosis, and Survival in High-Level Microsatellite Instability Sporadic Colorectal Cancer 1. *Clinical Cancer Research*, *7*, 2347–2356.

Mills, Anne M., Gradecki, Sarah E., Horton, Bethany J., Blackwell, Rebecca, Moskaluk, Christopher A., Mandell, James W., Mills, Stacey E., Cathro, Helen P. (2018). Diagnostic Efficiency in Digital Pathology. *American Journal of Surgical Pathology*, *42*(1), 53–59.

Mitmaker, Benjamin, Begin, Louis R., Gordon, Philip H. (1991). *Nuclear Shape as a Prognostic Discriminant in Colorectal Carcinoma*.

Mohammed, Z. M. A., Going, J. J., Edwards, J., Elsberger, B., Mcmillan, D. C. (2013). The relationship between lymphocyte subsets and clinico-pathological determinants of survival in patients with primary operable invasive ductal breast cancer. *British Journal of Cancer 2013 109:6*, *109*(6), 1676–1684.

Mohammed, Z. M. A., McMillan, D. C., Elsberger, B., Going, J. J., Orange, C., Mallon, E., Doughty, J. C., Edwards, J. (2012). Comparison of Visual and automated assessment of Ki-67 proliferative activity and their impact on outcome in primary operable invasive ductal breast cancer. *British Journal of Cancer 2012 106:2*, *106*(2), 383–388.

Morris, Eva J. A., Maughan, Nicola J., Forman, David, Quirke, Philip. (2007). Who to treat with adjuvant therapy in Dukes B/stage II colorectal cancer? The need for high quality pathology. *Gut*, *56*(10), 1419–1425.

Morson, Basil C. (Basil Clifford), Sobin, L. H. (1976). *Histological typing of intestinal tumours / B. C. Morson, in collaboration with L. H. Sobin and pathologists in seventeen countries*.

Murphy, Kathleen M., Zhang, Shengle, Geiger, Tanya, Hafez, Michael J., Bacher, Jeff, Berg, Karin D., Eshleman, James R. (2006). Comparison of the Microsatellite Instability Analysis System and the Bethesda Panel for the Determination of Microsatellite Instability in Colorectal Cancers. *The Journal of Molecular Diagnostics : JMD*, *8*(3), 305.

Nagtegaal, Iris D., Marijnen, Corrie A. M., Kranenbarg, Elma Klein, Mulder-Stapel, Adri, Hermans, Jo, van de Velde, Cornelis J. H., van Krieken, J. Han J. M. (2001). Local and distant recurrences in rectal cancer patients are predicted by the nonspecific immune response; specific immune response has only a systemic effect - A histopathological and immunohistochemical study. *BMC Cancer*, *1*(1), 1–11.

Naito, Yoshitaka, Saito, Kazuya, Shiiba, Kenichi, Ohuchi, Akio, Saigenji, Katsunori, Nagura, Hiroshi, Ontani1, Haruo. (1998). CD8+ T Cells Infiltrated within Cancer Cell Nests as a Prognostic Factor in Human Colorectal Cancer. *CANCER RESEARCH*, *58*, 3491–3494.

Najdi, Rani, Holcombe, Randall, Waterman, Marian. (2011). Wnt signaling and colon carcinogenesis: Beyond APC. *Journal of Carcinogenesis*, *10*.

Nakagawa, Hidewaki, Liyanarachchi, Sandya, Davuluri, Ramana V., Auer, Herbert, Martin, Edward W., De La Chapelle, Albert, Frankel, Wendy L. (2004). Role of cancer-associated stromal fibroblasts in metastatic colon cancer to the liver and their expression profiles. *Oncogene 2004 23:44*, *23*(44), 7366–7377.

Nash, Garrett M., Gimbel, Mark, Shia, Jinru, Nathanson, Daniel R., Ndubuisi, MacKevin I., Zeng, Zhao Shi, Kemeny, Nancy, Paty, Philip B. (2010a). KRAS mutation correlates with accelerated metastatic progression in patients with colorectal liver metastases. *Annals of Surgical Oncology*, *17*(2), 572–578.

Nash, Garrett M., Gimbel, Mark, Shia, Jinru, Nathanson, Daniel R., Ndubuisi, MacKevin I., Zeng, Zhao Shi, Kemeny, Nancy, Paty, Philip B. (2010b). KRAS mutation correlates with accelerated metastatic progression in patients with colorectal liver metastases. *Annals of Surgical Oncology*, *17*(2), 572–578.

Nassar, Aziza, Cohen, Cynthia, Agersborg, Sally S., Zhou, Weidong, Lynch, Kathleen A., Heyman, Eugene R., Olson, Allen, Lange, Holger, Siddiqui, Momin T. (2011). A new immunohistochemical ER/PR image analysis system: A multisite performance study. *Applied Immunohistochemistry and Molecular Morphology*, *19*(3), 195–202.

Nawa, Toru, Kato, Jun, Kawamoto, Hirofumi, Okada, Hiroyuki, Yamamoto, Hiroshi, Kohno, Hiroyuki, Endo, Hisayuki, Shiratori, Yasushi. (2008). Differences between right- and left-sided colon cancer in patient characteristics, cancer morphology and histology. *Journal of Gastroenterology and Hepatology*, *23*(3), 418–423.

Nayak, Jhasaketan, Mohanty, Pranita, Lenka, Anasuya, Sahoo, Nibedita, Agrawala, Sunil, Panigrahi, Sandeep. (2021). Histopathological and Immunohistochemical Evaluation of CDX2 and Ki67 in Colorectal Lesions with their Expression Pattern in Different Histologic Variants, Grade, and Stage of Colorectal Carcinomas. *Journal of Microscopy and Ultrastructure*, *9*(4), 183.

Nearchou, Ines P., Kajiwara, Yoshiki, Mochizuki, Satsuki, Harrison, David J., Caie, Peter D., Ueno, Hideki. (2019). Novel Internationally Verified Method Reports Desmoplastic Reaction as the Most Significant Prognostic Feature For Disease-specific Survival in Stage II Colorectal Cancer. *The American Journal of Surgical Pathology*, *43*(9), 1239–1248.

Nearchou, Ines P., Lillard, Kate, Gavriel, Christos G., Ueno, Hideki, Harrison, David J., Caie, Peter D. (2019a). Automated analysis of lymphocytic infiltration, tumor budding, and

their spatial relationship improves prognostic accuracy in colorectal cancer. *Cancer Immunology Research*, *7*(4), 609–620.

Nearchou, Ines P., Lillard, Kate, Gavriel, Christos G., Ueno, Hideki, Harrison, David J., Caie, Peter D. (2019b). Automated analysis of lymphocytic infiltration, tumor budding, and their spatial relationship improves prognostic accuracy in colorectal cancer. *Cancer Immunology Research*, *7*(4), 609–620.

Nearchou, Ines P., Lillard, Kate, Gavriel, Christos G., Ueno, Hideki, Harrison, David J., Caie, Peter D. (2019c). Automated analysis of lymphocytic infiltration, tumor budding, and their spatial relationship improves prognostic accuracy in colorectal cancer. *Cancer Immunology Research*, *7*(4), 609–620.

Nearchou, Ines P., Ueno, Hideki, Kajiwara, Yoshiki, Lillard, Kate, Mochizuki, Satsuki, Takeuchi, Kengo, Harrison, David J., Caie, Peter D. (2021). Automated Detection and Classification of Desmoplastic Reaction at the Colorectal Tumour Front Using Deep Learning. *Cancers 2021, Vol. 13, Page 1615*, *13*(7), 1615.

Nearchou, Ines, Soutar, Daniel, Ueno, Hideki, Harrison, David, Arandjelovic, Ognjen, Caie, Peter. (2021). A Comparison of Methods for Studying the Tumor Microenvironment's Spatial Heterogeneity in Digital Pathology Specimens. *Journal of Pathology Informatics*, *12*(1).

Nguyen, Huu Giao, Lundström, Oxana, Blank, Annika, Dawson, Heather, Lugli, Alessandro, Anisimova, Maria, Zlobec, Inti. (2021). Image-based assessment of extracellular mucin-to-tumor area predicts consensus molecular subtypes (CMS) in colorectal cancer. *Modern Pathology 2021 35:2*, *35*(2), 240–248.

Nielsen, Maartje, Morreau, Hans, Vasen, Hans F. A., Hes, Frederik J. (2011). MUTYH-associated polyposis (MAP). *Critical Reviews in Oncology/Hematology*, *79*(1), 1–16.

Nielsen, Patricia S., Riber-Hansen, Rikke, Jensen, Trine O., Schmidt, Henrik, Steiniche, Torben. (2013). Proliferation indices of phosphohistone H3 and Ki67: Strong prognostic markers in a consecutive cohort with stage I/II melanoma. *Modern Pathology*, *26*(3), 404–413.

Norton, Samuel E., Ward-Hartstonge, Kirsten A., Taylor, Edward S., Kemp, Roslyn A. (2015). Immune cell interplay in colorectal cancer prognosis. *World Journal of Gastrointestinal Oncology*, *7*(10), 221.

Öberg, Å., Samii, S., Stenling, R., Lindmark, G. (2001). Different occurrence of CD8+, CD45R0+, and CD68+ immune cells in regional lymph node metastases from colorectal cancer as potential prognostic predictors. *International Journal of Colorectal Disease 2001 17:1*, *17*(1), 25–29.

O'Brien, Michael J., Yang, Shi, Mack, Charline, Xu, Huihong, Huang, Christopher S., Mulcahy, Elizabeth, Amorosino, Mark, Farraye, Francis A. (2006). Comparison of microsatellite instability, CpG island methylation phenotype, BRAF and KRAS status in serrated polyps and traditional adenomas indicates separate pathways to distinct colorectal carcinoma end points. *American Journal of Surgical Pathology*, *30*(12), 1491–1501.

O'Connell, Jessica B., Maggard, Melinda A., Ko, Clifford Y. (2004). Colon Cancer Survival Rates With the New American Joint Committee on Cancer Sixth Edition Staging. *JNCI: Journal of the National Cancer Institute*, *96*(19), 1420–1425.

O'Connor, Kate, Li-Chang, Hector H., Kalloger, Steven E., Peixoto, Renata D., Webber, Douglas L., Owen, David A., Driman, David K., Kirsch, Richard, Serra, Stefano, Scudamore, Charles H., Renouf, Daniel J., Schaeffer, David F. (2015). Tumor budding is an independent adverse prognostic factor in pancreatic ductal adenocarcinoma. *American Journal of Surgical Pathology*, *39*(4), 472–478.

Okita, Akira, Takahashi, Shin, Ouchi, Kota, Inoue, Masahiro, Watanabe, Mika, Endo, Mareyuki, Honda, Hiroshi, Yamada, Yasuhide, Ishioka, Chikashi. (2018). Consensus molecular subtypes classification of colorectal cancer as a predictive factor for chemotherapeutic efficacy against metastatic colorectal cancer. *Oncotarget*, *9*(27), 18698.

Pai, Reetesh K., Hartman, Douglas, Schaeffer, David F., Rosty, Christophe, Shivji, Sameer, Kirsch, Richard, Pai, Rish K. (2021). Development and initial validation of a deep learning algorithm to quantify histological features in colorectal carcinoma including tumour budding/poorly differentiated clusters. *Histopathology*, *79*(3), 391–405.

Pai, Rish K., Bettington, Mark, Srivastava, Amitabh, Rosty, Christophe. (2019). An update on the morphology and molecular pathology of serrated colorectal polyps and associated carcinomas. *Modern Pathology 2019 32:10*, *32*(10), 1390–1415.

Palmqvist, R., Sellberg, P., Öberg, Å., Tavelin, B., Rutegård, Jn, Stenling, R. (1999). Low tumour cell proliferation at the invasive margin is associated with a poor prognosis in Dukes' stage B colorectal cancers. *British Journal of Cancer*, *79*(3–4), 577–581.

Pantanowitz L, Sinard JH, Henricks WH. (2013). *Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center* (Vol. 137, Issue (12), pp. 1710–1722).

Park, J. H., McMillan, D. C., Edwards, J., Horgan, P. G., Roxburgh, C. S. D. (2016). Comparison of the prognostic value of measures of the tumor inflammatory cell infiltrate and tumor-associated stroma in patients with primary operable colorectal cancer. *OncoImmunology*, *5*(3).

Park, J. H., Richards, C. H., McMillan, D. C., Horgan, P. G., Roxburgh, C. S. D. (2014a). The relationship between tumour stroma percentage, the tumour microenvironment and survival in patients with primary operable colorectal cancer. *Annals of Oncology*, *25*(3), 644–651.

Park, J. H., Richards, C. H., McMillan, D. C., Horgan, P. G., Roxburgh, C. S. D. (2014b). The relationship between tumour stroma percentage, the tumour microenvironment and survival in patients with primary operable colorectal cancer. *Annals of Oncology*, *25*(3), 644–651.

Park, J. H., Van Wyk, H., Roxburgh, C. S. D., Horgan, P. G., Edwards, J., McMillan, D. C. (2017). Tumour invasiveness, the local and systemic environment and the basis of staging systems in colorectal cancer. *British Journal of Cancer 2017 116:11*, *116*(11), 1444–1450.

Park, James H., Fuglestad, Anniken J., Køstner, Anne H., Oliwa, Agata, Graham, Janet, Horgan, Paul G., Roxburgh, Campbell S. D., Kersten, Christian, McMillan, Donald C. (2020). Systemic Inflammation and Outcome in 2295 Patients with Stage I–III Colorectal Cancer from Scotland and Norway: First Results from the ScotScan Colorectal Cancer Group. *Annals of Surgical Oncology*, *27*(8), 2784–2794.

Park, James H., McMillan, Donald C., Powell, Arfon G., Richards, Colin H., Horgan, Paul G., Edwards, Joanne, Roxburgh, Campbell S. D. (2015a). Evaluation of a tumor microenvironment-based prognostic score in primary operable colorectal cancer. *Clinical Cancer Research*, *21*(4), 882–888.

Park, James H., McMillan, Donald C., Powell, Arfon G., Richards, Colin H., Horgan, Paul G., Edwards, Joanne, Roxburgh, Campbell S. D. (2015b). Evaluation of a tumor microenvironment-based prognostic score in primary operable colorectal cancer. *Clinical Cancer Research*, *21*(4), 882–888.

Park, James H., Powell, Arfon G., Roxburgh, Campbell S. D., Horgan, Paul G., McMillan, Donald C., Edwards, Joanne. (2016). Mismatch repair status in patients with primary operable colorectal cancer: associations with the local and systemic tumour environment. *British Journal of Cancer 2016 114:5*, *114*(5), 562–570.

Park, James H., Watt, David G., Roxburgh, Campbell S. D., Horgan, Paul G., McMillan, Donald C. (2016). Colorectal cancer, systemic inflammation, and outcome: Staging the tumor and staging the host. *Annals of Surgery*, *263*(2), 326–336.

Patel, Meera, McSorley, Stephen T., Park, James H., Roxburgh, Campbell S. D., Edwards, Joann, Horgan, Paul G., McMillan, Donald C. (2018). The relationship between right-sided tumour location, tumour microenvironment, systemic inflammation, adjuvant therapy and survival in patients undergoing surgery for colon and rectal cancer. *British Journal of Cancer*, *118*(5), 705–712.

Pell, Robert, Oien, Karin, Robinson, Max, Pitman, Helen, Rajpoot, Nasir, Rittscher, Jens, Snead, David, Verrill, Clare, Driskell, Owen J., Hall, Andy, James, Jacqueline, Jones, Louise J., Craig, Clare, Sloan, Philip, Thomas, Gareth J., Elliott, Philip, Cheang, Maggie, Rodriguez-Justo, Manuel, Rees, Gabrielle, Salto-Tellez, Manuel, West, Nicholas P., Mirabile, Ilaria, Howlett, Emily, Stevenson, Laura, da Silva, Maria, Hartridge-Lambert, Sidonie, Beecham, Joseph M., Traub, Stephanie, Katugampola, Sidath, Blagden, Sarah, Morden, James. (2019). The use of digital pathology and image analysis in clinical trials. *The Journal of Pathology: Clinical Research*, *5*(2), 81–90.

Petersen, V. C., Baxter, K. J., Love, S. B., Shepherd, Neil A. (2002). Identification of objective pathological prognostic determinants and models of prognosis in Dukes' B colon cancer. *Gut*, *51*(1), 65.

Popat, Sanjay, Hubner, R., Houlston, R. S. (2005a). Systematic review of microsatellite instability and colorectal cancer prognosis. *Journal of Clinical Oncology*, *23*(3), 609–618.

Popat, Sanjay, Hubner, R., Houlston, R. S. (2005b). Systematic review of microsatellite instability and colorectal cancer prognosis. *Journal of Clinical Oncology*, *23*(3), 609–618.

Qaiser, Talha, Mukherjee, Abhik, Reddy PB, Chaitanya, Munugoti, Sai D., Tallam, Vamsi, Pitkäaho, Tomi, Lehtimäki, Taina, Naughton, Thomas, Berseth, Matt, Pedraza, Aníbal, Mukundan, Ramakrishnan, Smith, Matthew, Bhalerao, Abhir, Rodner, Erik, Simon, Marcel, Denzler, Joachim, Huang, Chao Hui, Bueno, Gloria, Snead, David, Ellis, Ian O., Ilyas, Mohammad, Rajpoot, Nasir. (2018). HER2 challenge contest: a detailed assessment of automated HER2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology*, *72*(2), 227–238.

Rau, Tilman T., Bettschen, Eva, Büchi, Carol, Christe, Lucine, Rohner, Amanda, Müller, Michael D., Carlson, Joseph W., Imboden, Sara, Zlobec, Inti. (2020). Prognostic impact of tumor budding in endometrial carcinoma within distinct molecular subgroups. *Modern Pathology 2020 34:1*, *34*(1), 222–232.

Raymond, Victoria M., Mukherjee, Bhramar, Wang, Fei, Huang, Shu Chen, Stoffel, Elena M., Kastrinos, Fay, Syngal, Sapna, Cooney, Kathleen A., Gruber, Stephen B. (2013). Elevated Risk of Prostate Cancer Among Men With Lynch Syndrome. *Journal of Clinical Oncology*, *31*(14), 1713.

Reimers, M. S., Zeestraten, E. C. M., Van Alphen, T. C., Dekker, J. W. T., Putter, H., Saadatmand, S., Liefers, G. J., Van De Velde, C. J. H., Kuppen, P. J. K. (2014). Combined analysis of biomarkers of proliferation and apoptosis in colon cancer: An immunohistochemistry-based study using tissue microarray. *International Journal of Colorectal Disease*, *29*(9), 1043–1052.

Richards, C. H., Flegg, K. M., Roxburgh, C. S. D., Going, J. J., Mohammed, Z., Horgan, P. G., McMillan, D. C. (2012). The relationships between cellular components of the peritumoural inflammatory response, clinicopathological characteristics and survival in patients with primary operable colorectal cancer. *British Journal of Cancer*, *106*(12), 2010–2015.

Richards, Colin H., Roxburgh, Campbell S. D., Powell, Arfon G., Foulis, Alan K., Horgan, Paul G., McMillan, Donald C. (2014). The clinical utility of the local inflammatory response in colorectal cancer. *European Journal of Cancer*, *50*(2), 309–319.

Richman, Susan. (2015). Deficient mismatch repair: Read all about it (Review). *International Journal of Oncology*, *47*(4), 1189.

Rmali, K. A., Puntis, M. C. A., Jiang, W. G. (2007). Tumour-associated angiogenesis in human colorectal cancer. *Colorectal Disease*, *9*(1), 3–14.

Roepman, Paul, Schlicker, Andreas, Tabernero, Josep, Majewski, Ian, Tian, Sun, Moreno, Victor, Snel, Mireille H., Chresta, Christine M., Rosenberg, Robert, Nitsche, Ulrich, Macarulla, Teresa, Capella, Gabriel, Salazar, Ramon, Orphanides, George, Wessels, Lodewyk F. A., Bernards, Rene, Simon, Iris M. (2014a). Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. *International Journal of Cancer*, *134*(3), 552–562.

Roepman, Paul, Schlicker, Andreas, Tabernero, Josep, Majewski, Ian, Tian, Sun, Moreno, Victor, Snel, Mireille H., Chresta, Christine M., Rosenberg, Robert, Nitsche, Ulrich, Macarulla, Teresa, Capella, Gabriel, Salazar, Ramon, Orphanides, George, Wessels, Lodewyk F. A., Bernards, Rene, Simon, Iris M. (2014b). Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. *International Journal of Cancer*, *134*(3), 552–562.

Romo-Bucheli, David, Janowczyk, Andrew, Gilmore, Hannah, Romero, Eduardo, Madabhushi, Anant. (2017). A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in estrogen receptor positive breast cancers. *Cytometry Part A*, *91*(6), 566–573.

Ronneberger, Olaf, Fischer, Philipp, Brox, Thomas. (2015a). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9351*, 234–241.

Ronneberger, Olaf, Fischer, Philipp, Brox, Thomas. (2015b). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9351*, 234–241.

Ronneberger, Olaf, Fischer, Philipp, Brox, Thomas. (2015c). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9351*, 234–241.

Ropponen, K., Tammi, M., Parkkinen, J., Eskelinen, M., Tammi, R., Lipponen, P., Ågren, U., Alhava, E., Kosma, V. (1998). Tumor cell-associated hyaluronan as an unfavorable prognostic factor in colorectal cancer. *Cancer Research*.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408.

Roseweir, Antonia K., McMillan, Donald C., Horgan, Paul G., Edwards, Joanne. (2017). Colorectal cancer subtypes: Translation to routine clinical pathology. In *Cancer Treatment Reviews* (Vol. 57, pp. 1–7). W.B. Saunders Ltd.

Roseweir, Antonia K., Park, James H., Hoorn, Sanne ten, Powell, Arfon G. M. T., Aherne, Susan, Roxburgh, Campbell S. D., McMillan, Donald C., Horgan, Paul G., Ryan, Elizabeth, Sheahan, Kieran, Vermeulen, Louis, Paul, James, Harkin, Andrea, Graham, Janet, Sansom, Owen, Church, David N., Tomlinson, Ian, Saunders, Mark, Iveson, Tim J., Edwards, Joanne. (2020). Histological phenotypic subtypes predict recurrence risk and response to adjuvant chemotherapy in patients with stage III colorectal cancer. *Journal of Pathology: Clinical Research*, *6*(4), 283–296.

Roseweir, Antonia K., Park, James Hugh, ten Hoorn, Sanne, Powell, Arfon G. M. T., Roxburgh, Campbell S. D., McMillan, Donald C., Horgan, Paul G., Vermeulen, Louis, Edwards, Joanne. (2018a). Phenotypic subtypes as a novel validated prognostic classification system for patients with colorectal cancer. *Journal of Clinical Oncology*, *36*(4_suppl), 625.

Roseweir, Antonia K., Park, James Hugh, ten Hoorn, Sanne, Powell, Arfon GMT, Roxburgh, Campbell SD, McMillan, Donald C., Horgan, Paul G., Vermeulen, Louis, Edwards, Joanne. (2018b). Phenotypic subtypes as a novel validated prognostic classification system for patients with colorectal cancer. *Journal of Clinical Oncology*, *36*(4_suppl), 625–625.

Rowan, A. J., Lamlum, H., Ilyas, M., Wheeler, J., Straub, J., Papadopoulou, A., Bicknell, D., Bodmer, W. F., Tomlinson, I. P. M. (2000). APC mutations in sporadic colorectal tumors: A mutational 'hotspot' and interdependence of the 'two hits'. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(7), 3352–3357.

Roxburgh, C. S. D., McMillan, D. C. (2012). The role of the in situ local inflammatory response in predicting recurrence and survival in patients with primary operable colorectal cancer. *Cancer Treatment Reviews*, *38*(5), 451–466.

Roxburgh, C. S., Richards, C. H., Macdonald, A. I., Powell, A. G., McGlynn, L. M., McMillan, D. C., Horgan, P. G., Edwards, J., Shiels, P. G. (2013). The in situ local immune response, tumour senescence and proliferation in colorectal cancer. *British Journal of Cancer 2013 109:8*, *109*(8), 2207–2216.

Roxburgh, Campbell S. D., McMillan, Donald C., Anderson, John H., McKee, Ruth F., Horgan, Paul G., Foulis, Alan K. (2010). Elastica staining for venous invasion results in superior prediction of cancer-specific survival in colorectal cancer. *Annals of Surgery*, *252*(6), 989–997.

Roxburgh, Campbell S. D., Mcmillan, Donald C., Richards, Colin H., Atwan, Manal, Anderson, John H., Harvey, Tim, Horgan, Paul G., Foulis, Alan K. (2014). The clinical utility of the combination of T stage and venous invasion to predict survival in patients undergoing surgery for colorectal cancer. *Annals of Surgery*, *259*(6), 1156–1165.

Roxburgh, Campbell S. D., Salmond, Jonathan M., Horgan, Paul G., Oien, Karin A., McMillan, Donald C. (2009a). Comparison of the prognostic value of inflammation-based pathologic and biochemical criteria in patients undergoing potentially curative resection for colorectal cancer. *Annals of Surgery*, *249*(5), 788–793.

Roxburgh, Campbell S. D., Salmond, Jonathan M., Horgan, Paul G., Oien, Karin A., McMillan, Donald C. (2009b). Tumour inflammatory infiltrate predicts survival following curative resection for node-negative colorectal cancer. *European Journal of Cancer*, *45*(12), 2138–2145.

Ryan, N. A. J., Glaire, M. A., Blake, D., Cabrera-Dandy, M., Evans, D. G., Crosbie, E. J. (2019). The proportion of endometrial cancers associated with Lynch syndrome: a systematic review of the literature and meta-analysis. *Genetics in Medicine*, *21*(10), 2167–2180.

Sadanandam, Anguraj, Lyssiotis, Costas A., Homicsko, Krisztian, Collisson, Eric A., Gibb, William J., Wullschleger, Stephan, Ostos, Liliane C. Gonzalez, Lannon, William A., Grotzinger, Carsten, Del Rio, Maguy, Lhermitte, Benoit, Olshen, Adam B., Wiedenmann, Bertram, Cantley, Lewis C., Gray, Joe W., Hanahan, Douglas. (2013). A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nature Medicine 2013 19:5*, *19*(5), 619–625.

Sadler, T. W. (2018). *Langman's medical embryology*.

Saltz, Joel, Gupta, Rajarsi, Hou, Le, Kurc, Tahsin, Singh, Pankaj, Nguyen, Vu, Samaras, Dimitris, Shroyer, Kenneth R., Zhao, Tianhao, Batiste, Rebecca, Van Arnam, John, Caesar-Johnson, Samantha J., Demchok, John A., Felau, Ina, Kasapi, Melpomeni, Ferguson, Martin L., Hutter, Carolyn M., Sofia, Heidi J., Tarnuzzer, Roy, Wang, Zhining, Yang, Liming, Zenklusen, Jean C., … Thorsson, Vésteinn. (2018). Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell Reports*, *23*(1), 181-193.e7.

Schlicker, Andreas, Beran, Garry, Chresta, Christine M., McWalter, Gael, Pritchard, Alison, Weston, Susie, Runswick, Sarah, Davenport, Sara, Heathcote, Kerry, Castro, Denis Alferez, Orphanides, George, French, Tim, Wessels, Lodewyk F. A. (2012). Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines. *BMC Medical Genomics*, *5*(1), 1–15.

Schmaus, Anja, Bauer, Jochen, Sleeman, Jonathan P. (2014). Sugars in the microenvironment: the sticky problem of HA turnover in tumors. *Cancer and Metastasis Reviews*, *33*(4), 1059–1079.

Schwab, Ulrich, Stein, Harald, Gerdes, Johannes, Lemke, Hilmar, Kirchner, Hartmut, Schaadt, Michael, Diehl, Volker. (1982). Production of a monoclonal antibody specific for Hodgkin and Sternberg-Reed cells of Hodgkin's disease and a subset of normal lymphoid cells. *Nature*, *299*(5878), 65–67.

Selvaraju, Ramprasaath R., Cogswell, Michael, Das, Abhishek, Vedantam, Ramakrishna, Parikh, Devi, Batra, Dhruv. (2016). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision*, *128*(2), 336–359.

Shaban, Muhammad, Khurram, Syed Ali, Fraz, Muhammad Moazam, Alsubaie, Najah, Masood, Iqra, Mushtaq, Sajid, Hassan, Mariam, Loya, Asif, Rajpoot, Nasir M. (2019). A Novel Digital Score for Abundance of Tumour Infiltrating Lymphocytes Predicts Disease Free Survival in Oral Squamous Cell Carcinoma. *Scientific Reports 2019 9:1*, *9*(1), 1–13.

Shelhamer, Evan, Long, Jonathan, Darrell, Trevor. (2017). Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(4), 640–651.

Shia, Jinru, Wang, Hangjun, Nash, Garrett M., Klimstra, David S. (2012). Lymph node staging in colorectal cancer: Revisiting the benchmark of at least 12 lymph nodes in R0 resection. *Journal of the American College of Surgeons*, *214*(3), 348–355.

Shin, Il Yong, Sung, Na Young, Lee, Youn Soo, Kwon, Taek Soo, Si, Yoon, Lee, Yoon Suk, Oh, Seong Taek, Lee, In Kyu. (2014). The Expression of Multiple Proteins as Prognostic Factors in Colorectal Cancer: Cathepsin D, p53, COX-2, Epidermal Growth Factor Receptor, C-erbB-2, and Ki-67. *Gut and Liver*, *8*(1), 13.

Simonyan, Karen, Zisserman, Andrew. (2015, September 4). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Sirinukunwattana, Korsuk, Domingo, Enric, Richman, Susan D., Redmond, Keara L., Blake, Andrew, Verrill, Clare, Leedham, Simon J., Chatzipli, Aikaterini, Hardy, Claire, Whalley, Celina M., Wu, Chieh-hsi, Beggs, Andrew D., McDermott, Ultan, Dunne, Philip D., Meade, Angela, Walker, Steven M., Murray, Graeme I., Samuel, Leslie, Seymour, Matthew, Tomlinson, Ian, Quirke, Phil, Maughan, Timothy, Rittscher, Jens, Koelzer, Viktor H., Viktor Koelzer, Professor H., Jens Rittscher, Professor. (2021). Artificial intelligence/machine learning Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut*, *70*, 544–554.

Sirinukunwattana, Korsuk, Domingo, Enric, Richman, Susan D., Redmond, Keara L., Blake, Andrew, Verrill, Clare, Leedham, Simon J., Chatzipli, Aikaterini, Hardy, Claire, Whalley, Celina M., Wu, Chieh-hsi, Beggs, Andrew D., McDermott, Ultan, Dunne, Philip D., Meade, Angela, Walker, Steven M., Murray, Graeme I., Samuel, Leslie, Seymour, Matthew, Tomlinson, Ian, Quirke, Phil, Maughan, Timothy, Rittscher, Jens, Koelzer, Viktor H. (2021). Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut*, *70*(3), 544–554.

Skrede, Ole Johan, De Raedt, Sepp, Kleppe, Andreas, Hveem, Tarjei S., Liestøl, Knut, Maddison, John, Askautrud, Hanne A., Pradhan, Manohar, Nesheim, John Arne, Albregtsen, Fritz, Farstad, Inger Nina, Domingo, Enric, Church, David N., Nesbakken, Arild, Shepherd, Neil A., Tomlinson, Ian, Kerr, Rachel, Novelli, Marco, Kerr, David J., Danielsen, Håvard E. (2020). Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *The Lancet*, *395*(10221), 350–360.

Smit, Marloes A., van Pelt, Gabi W., Terpstra, Valeska, Putter, Hein, Tollenaar, Rob A. E. M., Mesker, Wilma E., van Krieken, J. Han J. M. (2021a). Tumour-stroma ratio

outperforms tumour budding as biomarker in colon cancer: a cohort study. *International Journal of Colorectal Disease*, *36*(12), 2729–2737.

Smit, Marloes A., van Pelt, Gabi W., Terpstra, Valeska, Putter, Hein, Tollenaar, Rob A. E. M., Mesker, Wilma E., van Krieken, J. Han J. M. (2021b). Tumour-stroma ratio outperforms tumour budding as biomarker in colon cancer: a cohort study. *International Journal of Colorectal Disease*, *36*(12), 2729–2737.

Smit, Marloes, Van Pelt, Gabi, Roodvoets, Annet, Kranenbarg, Elma Meershoek Klein, Putter, Hein, Tollenaar, Rob, Van Krieken, J. Han, Mesker, Wilma. (2019). Uniform Noting for International Application of the Tumor-Stroma Ratio as an Easy Diagnostic Tool: Protocol for a Multicenter Prospective Cohort Study. *JMIR Research Protocols*, *8*(6).

Snead, David R. J., Tsang, Yee Wah, Meskiri, Aisha, Kimani, Peter K., Crossman, Richard, Rajpoot, Nasir M., Blessing, Elaine, Chen, Klaus, Gopalakrishnan, Kishore, Matthews, Paul, Momtahan, Navid, Read-Jones, Sarah, Sah, Shatrughan, Simmons, Emma, Sinha, Bidisa, Suortamo, Sari, Yeo, Yen, El Daly, Hesham, Cree, Ian A. (2016). Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology*, *68*(7), 1063–1072.

Sobin, Leslie H., Hermanek, Paul, Hutter, Robert V. P. (1988). TNM classification of malignant tumors. A comparison between the new (1987) and the old editions. *Cancer*, *61*(11), 2310–2314.

Spratt, John S., Spjut, Harlan J. (1967). PREVALENCE AND PROGNOSIS OF INDIVIDUAL CLINICAL AND PATHOLOGIC VARIABLES ASSOCIATED WITH COLORECTAL CARCINOMA. *Cancer*, *20*(11), 1976–1985.

Spriggs, A. I. (1969). Automatic scanning for cervical smears. *Journal of Clinical Pathology*, *S2-3*(1), 1–6.

Stålhammar, Gustav, Robertson, Stephanie, Wedlund, Lena, Lippert, Michael, Rantalainen, Mattias, Bergh, Jonas, Hartman, Johan. (2018). Digital image analysis of Ki67 in hot spots is superior to both manual Ki67 and mitotic counts in breast cancer. *Histopathology*, *72*(6), 974–989.

Steiner, David F., Macdonald, Robert, Liu, Yun, Truszkowski, Peter, Hipp, Jason D., Gammage, Christopher, Thng, Florence, Peng, Lily, Stumpe, Martin C. (2018). Impact

of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. *The American Journal of Surgical Pathology*, *42*(12), 1636.

Stenman, Sebastian, Bychkov, Dmitrii, Kucukel, Hakan, Linder, Nina, Haglund, Caj, Arola, Johanna, Lundin, Johan. (2021). Antibody Supervised Training of a Deep Learning Based Algorithm for Leukocyte Segmentation in Papillary Thyroid Carcinoma. *IEEE Journal of Biomedical and Health Informatics*, *25*(2), 422–428.

Stewart, C. J. R., Morris, M., De Boer, B., Iacopetta, B. (2007). Identification of serosal invasion and extramural venous invasion on review of Dukes' stage B colonic carcinomas and correlation with survival. *Histopathology*, *51*(3), 372–378.

Stoffel, Elena M., Mercado, Rowena C., Kohlmann, Wendy, Ford, Beth, Grover, Shilpa, Conrad, Peggy, Blanco, Amie, Shannon, Kristen M., Powell, Mark, Chung, Daniel C., Terdiman, Jonathan, Gruber, Stephen B., Syngal, Sapna. (2010). Prevalence and Predictors of Appropriate Colorectal Cancer Surveillance in Lynch Syndrome. *The American Journal of Gastroenterology*, *105*(8), 1851.

Sueyama, Takahiro, Kajiwara, Yoshiki, Mochizuki, Satsuki, Shimazaki, Hideyuki, Shinto, Eiji, Hase, Kazuo, Ueno, Hideki. (2021). Periostin as a key molecule defining desmoplastic environment in colorectal cancer. *Virchows Archiv : An International Journal of Pathology*, *478*(5), 865–874.

Sung, Hyuna, Ferlay, Jacques, Siegel, Rebecca L., Laversanne, Mathieu, Soerjomataram, Isabelle, Jemal, Ahmedin, Bray, Freddie. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249.

Takagi, Sho, Kumagai, Shinji, Kinouchi, Yoshitaka, Hiwatashi, Nobuo, Nagashima, Fumio, Takahashi, Seiichi, Shimosegawa, Tooru. (2002). High Ki-67 labeling index in human colorectal cancer with microsatellite instability. *Anticancer Research*, *22*(6A), 3241–3244.

Talbot, I. C., Ritchie, SHEILA, Leighton, MONICA, Hughes, A. O., Bussey, H. J. R., Morson, B. C. (1981). Invasion of veins by carcinoma of rectum: method of detection, histological features and significance. *Histopathology*, *5*(2), 141–163.

Tawadros, Patrick S., Paquette, Ian M., Hanly, Ann M., Mellgren, Anders F., Rothenberger, David A., Madoff, Robert D. (2015). Adenocarcinoma of the rectum in patients under

age 40 is increasing: Impact of signet-ring cell histology. *Diseases of the Colon and Rectum*, *58*(5), 474–478.

Thakur, Nishant, Ailia, Muhammad Joan, Chong, Yosep, Shin, Ok Ran, Yim, Kwangil. (2022). Tumor Budding as a Marker for Poor Prognosis and Epithelial–Mesenchymal Transition in Lung Cancer: A Systematic Review and Meta-Analysis. *Frontiers in Oncology*, *12*, 828999.

Therkildsen, C., Ladelund, S., Rambech, E., Persson, A., Petersen, A., Nilbert, M. (2015). Glioblastomas, astrocytomas and oligodendrogliomas linked to Lynch syndrome. *European Journal of Neurology*, *22*(4), 717–724.

Thrall, Michael J., Wimmer, Jana L., Schwartz, Mary R. (2015). Validation of Multiple Whole Slide Imaging Scanners Based on the Guideline From the College of American Pathologists Pathology and Laboratory Quality Center. *Archives of Pathology & Laboratory Medicine*, *139*(5), 656–664.

Thynne, G. S., Weiland, L. H., Moertel, C. G., Silvers, A. (1980). Correlation of histopathologic characteristics of primary tumor and uninvolved regional lymph nodes in Dukes' class C colonic carcinoma with prognosis. *Mayo Clinic Proceedings*, *55*(4), 243–245.

Tian, Xiao, Azpurua, Jorge, Ke, Zhonghe, Augereau, Adeline, Zhang, Zhengdong D., Vijg, Jan, Gladyshev, Vadim N., Gorbunova, Vera, Seluanov, Andrei. (2015). INK4 locus of the tumor-resistant rodent, the naked mole rat, expresses a functional p15/p16 hybrid isoform. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(4), 1053–1058.

Tong, Guo Jun, Zhang, Gui Yang, Liu, Jian, Zheng, Zhao Zheng, Chen, Yan, Niu, Ping Ping, Xu, Xu Ting. (2018). Comparison of the eighth version of the American Joint Committee on Cancer manual to the seventh version for colorectal cancer: A retrospective review of our data. *World Journal of Clinical Oncology*, *9*(7), 148.

Torlakovic, Emina, Skovlund, Eva, Snover, Dale C., Torlakovic, Goran, Nesland, Jahn M. (2003). Morphologic reappraisal of serrated colorectal polyps. *The American Journal of Surgical Pathology*, *27*(1), 65–81.

Toyota, Minoru, Ahuja, Nita, Ohe-Toyota, Mutsumi, Herman, James G., Baylin, Stephen B., Issa, Jean Pierre J. (1999). CpG island methylator phenotype in colorectal cancer. *Proceedings of the National Academy of Sciences*, *96*(15), 8681–8686.

Trahearn, Nicholas, Tsang, Yee Wah, Cree, Ian A., Snead, David, Epstein, David, Rajpoot, Nasir. (2017). Simultaneous automatic scoring and co-registration of hormone receptors in tumor areas in whole slide images of breast cancer tissue slides. *Cytometry. Part A : The Journal of the International Society for Analytical Cytology*, *91*(6), 585–594.

Trinh, Anne, Trumpi, Kari, De Sousa E Melo, Felipe, Wang, Xin, De Jong, Joan H., Fessler, Evelyn, Kuppen, Peter J. K., Reimers, Marlies S., Swets, Marloes, Koopman, Miriam, Nagtegaal, Iris D., Jansen, Marnix, Hooijer, Gerrit K. J., Offerhaus, George J. A., Kranenburg, Onno, Punt, Cornelis J., Medema, Jan Paul, Markowetz, Florian, Vermeulen, Louis. (2017). Practical and robust identification of molecular subtypes in colorectal cancer by immunohistochemistry. *Clinical Cancer Research*, *23*(2), 387–398.

Turkki, Riku, Linder, Nina, Kovanen, Panu E., Pellinen, Teijo, Lundin, Johan. (2016). Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples. *Journal of Pathology Informatics*, *7*(1), 38.

Turnbull, R. B., Kyle, K., Watson, F. R., Spratt, J. (1967). Cancer of the colon: the influence of the no-touch isolation technic on survival rates. *Annals of Surgery*, *166*(3), 420.

Ueno, H., Hase, K., Mochizuki, H. (2002). Criteria for extramural perineural invasion as a prognostic factor in rectal cancer. *British Journal of Surgery*, *88*(7), 994–1000.

Ueno, H., Jones, A., Jass, J. R., Talbot, I. C. (2002a). Clinicopathological significance of the `keloid-like' collagen and myxoid stroma in advanced rectal cancer. *Histopathology*, *40*(4), 327–334.

Ueno, H., Jones, A., Jass, J. R., Talbot, I. C. (2002b). Clinicopathological significance of the 'keloid-like' collagen and myxoid stroma in advanced rectal cancer. *Histopathology*, *40*(4), 327–334.

Ueno, H., Jones, A. M., Wilkinson, K. H., Jass, J. R. (2004). Histological categorisation of fibrotic cancer stroma in advanced rectal cancer. *Gut*, *53*, 581–586.

Ueno, Hideki, Kanemitsu, Yukihide, Sekine, Shigeki, Ishiguro, Megumi, Ito, Eisaku, Hashiguchi, Yojiro, Kondo, Fukuo, Shimazaki, Hideyuki, Mochizuki, Satsuki, Kajiwara, Yoshiki, Shinto, Eiji, Yamamoto, Junji. (2017). Desmoplastic Pattern at the Tumor Front Defines Poor-prognosis Subtypes of Colorectal Cancer. *The American Journal of Surgical Pathology*, *41*(11), 1506–1512.

Ueno, Hideki, Shirouzu, Kazuo, Eishi, Yoshinobu, Yamada, Kazutaka, Kusumi, Takaya, Kushima, Ryoji, Ikegami, Masahiro, Murata, Akihiko, Okuno, Kiyotaka, Sato, Toshihiko, Ajioka, Yoichi, Ochiai, Atsushi, Shimazaki, Hideyuki, Nakamura, Takahiro, Kawachi, Hiroshi, Kojima, Motohiro, Akagi, Yoshito, Sugihara, Kenichi. (2013). Characterization of perineural invasion as a component of colorectal cancer staging. *The American Journal of Surgical Pathology*, *37*(10), 1542–1549.

Uguen, Arnaud, Uguen, Marie, Guibourg, Briac, Talagas, Matthieu, Marcorelles, Pascale, De Braekeleer, Marc. (2018). The p16-Ki-67-HMB45 Immunohistochemistry Scoring System is Highly Concordant with the Fluorescent in Situ Hybridization Test to Differentiate between Melanocytic Nevi and Melanomas. *Applied Immunohistochemistry and Molecular Morphology*, *26*(6), 361–367.

Uprety, Dipesh, Adjei, Alex A. (2020). KRAS: From undruggable to a druggable Cancer Target. *Cancer Treatment Reviews*, *89*, 102070.

Vabi, Benjamin W., Gibbs, John F., Parker, Glenn S. (2021). Implications of the growing incidence of global colorectal cancer. *Journal of Gastrointestinal Oncology*, *12*(Suppl 2), S387.

van Pelt, Gabi W., Hansen, Torben F., Bastiaannet, Esther, Kjær-Frifeldt, Sanne, Krieken, J. Han JM van, Tollenaar, Rob AEM, Sørensen, Flemming B., Mesker, Wilma E. (2016). Stroma-High Lymph Node Involvement Predicts Poor Survival More Accurately for Patients with Stage III Colon Cancer. *Journal of Medical & Surgical Pathology*, *1*(2), 1–8.

van Pelt, Gabi W., Sandberg, Tessa P., Morreau, Hans, Gelderblom, Hans, van Krieken, J. Han J. M., Tollenaar, Rob A. E. M., Mesker, Wilma E. (2018a). The tumour–stroma ratio in colon cancer: the biological role and its prognostic impact. *Histopathology*, *73*(2), 197–206.

van Pelt, Gabi W., Sandberg, Tessa P., Morreau, Hans, Gelderblom, Hans, van Krieken, J. Han J. M., Tollenaar, Rob A. E. M., Mesker, Wilma E. (2018b). The tumour–stroma ratio in colon cancer: the biological role and its prognostic impact. *Histopathology*, *73*(2), 197–206.

van Wyk, H. C., Going, James, Horgan, Paul, McMillan, Donald C. (2017). The role of perineural invasion in predicting survival in patients with primary operable colorectal cancer: A systematic review. *Critical Reviews in Oncology/Hematology*, *112*, 11–20.

van Wyk, H. C., Park, James, Roxburgh, Campbell, Horgan, Paul, Foulis, Alan, McMillan, Donald C. (2015). The role of tumour budding in predicting survival in patients with primary operable colorectal cancer: A systematic review. In *Cancer Treatment Reviews* (Vol. 41, Issue 2, pp. 151–159). W.B. Saunders Ltd.

van Wyk, Hester C., Roseweir, Antonia, Alexander, Peter, Park, James H., Horgan, Paul G., McMillan, Donald C., Edwards, Joanne. (2019). The Relationship Between Tumor Budding, Tumor Microenvironment, and Survival in Patients with Primary Operable Colorectal Cancer. *Annals of Surgical Oncology*, *26*(13), 4397–4404.

Vargas, Teodoro, Moreno-Rubio, Juan, Herranz, Jesús, Cejas, Paloma, Molina, Susana, González-Vallinas, Margarita, Ramos, Ricardo, Burgos, Emilio, Aguayo, Cristina, Custodio, Ana B., Reglero, Guillermo, Feliu, Jaime, Ramírez de Molina, Ana. (2014). Genes associated with metabolic syndrome predict disease-free survival in stage II colorectal cancer patients. A novel link between metabolic dysregulation and colorectal cancer. *Molecular Oncology*, *8*(8), 1469–1481.

Vass, D. G., Ainsworth, R., Anderson, J. H., Murray, D., Foulis, A. K. (2004). The value of an elastic tissue stain in detecting venous invasion in colorectal cancer. *Journal of Clinical Pathology*, *57*(7), 769.

Vayrynen, Juha P., Lau, Mai Chan, Haruki, Koichiro, Vayrynen, Sara A., Costa, Andressa Dias, Borowsky, Jennifer, Zhao, Melissa, Fujiyoshi, Kenji, Arima, Kota, Twombly, Tyler S., Kishikawa, Junko, Gu, Simeng, Aminmozaffari, Saina, Shi, Shanshan, Baba, Yoshifumi, Akimoto, Naohiko, Ugai, Tomotaka, Silva, Annacarolina Da, Song, Mingyang, Wu, Kana, Chan, Andrew T., Nishihara, Reiko, Fuchs, Charles S., Meyerhardt, Jeffrey A., Giannakis, Marios, Ogino, Shuji, Nowak, Jonathan A. (2020). Prognostic Significance of Immune Cell Populations Identified by Machine Learning in Colorectal Cancer Using Routine Hematoxylin and Eosin–Stained Sections. *Clinical Cancer Research*, *26*(16), 4326–4338.

Vermeulen, Louis, De Sousa E Melo, Felipe, Van Der Heijden, Maartje, Cameron, Kate, De Jong, Joan H., Borovski, Tijana, Tuynman, Jurriaan B., Todaro, Matilde, Merz, Christian, Rodermond, Hans, Sprick, Martin R., Kemper, Kristel, Richel, Dick J., Stassi, Giorgio, Medema, Jan Paul. (2010). Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nature Cell Biology 2010 12:5*, *12*(5), 468–476.

Vogelstein, B., Fearon, ER, Hamilton, SR, Kern, SE, Preisinger, AC, Leppert, M., Nakamura, Y., White, R., Smits, AM, Bos, JL. (1988). Genetic alterations during colorectal-tumor development. *The New England Journal of Medicine*, *319*(9), 525–532.

Waghray, Meghna, Yalamanchili, Malica, Magliano, Marina Pasca Di, Simeone, Diane M. (2013). Deciphering the Role of Stroma in Pancreatic Cancer. *Current Opinion in Gastroenterology*, *29*(5), 537.

Wang, Cassia B., Shahjehan, Faisal, Merchea, Amit, Li, Zhuo, Bekaii-Saab, Tanios S., Grothey, Axel, Colibaseanu, Dorin T., Kasi, Pashtoon M. (2019). Impact of tumor location and variables associated with overall survival in patients with colorectal cancer: A mayo clinic colon and rectal cancer registry study. *Frontiers in Oncology*, *9*(FEB), 76.

Wang, Xiangxue, Barrera, Cristian, Bera, Kaustav, Viswanathan, Vidya Sankar, Azarianpour-Esfahani, Sepideh, Koyuncu, Can, Velu, Priya, Feldman, Michael D., Yang, Michael, Fu, Pingfu, Schalper, Kurt A., Mahdi, Haider, Lu, Cheng, Velcheti, Vamsidhar, Madabhushi, Anant. (2022). Spatial interplay patterns of cancer nuclei and tumor-infiltrating lymphocytes (TILs) predict clinical benefit for immune checkpoint inhibitors. *Science Advances*, *8*(22), 3966.

Warschkow, Rene, Sulz, Michael C., Marti, Lukas, Tarantino, Ignazio, Schmied, Bruno M., Cerny, Thomas, Güller, Ulrich. (2016). Better survival in right-sided versus left-sided stage I - III colon cancer patients. *BMC Cancer*, *16*(1), 1–14.

Watson, Martin, Lea, Dordi, Skaland, Ivar, Hagland, Hanne R., Søreide, Kjetil. (2020). Loss of MSH3 Is Not Related to EMAST in Colorectal Cancer. Digitalised Automated Expression Analysis in a Population-Based Cohort. *European Journal of Surgical Oncology*, *46*(2), e108–e109.

Weinstein, Ronald S. (1986). Prospects for telepathology. (Editorial). *Human Path*, *17*(5), 433–434.

Weisenberger, Daniel J., Siegmund, Kimberly D., Campan, Mihaela, Young, Joanne, Long, Tiffany I., Faasse, Mark A., Kang, Gyeong Hoon, Widschwendter, Martin, Weener, Deborah, Buchanan, Daniel, Koh, Hoey, Simms, Lisa, Barker, Melissa, Leggett, Barbara, Levine, Joan, Kim, Myungjin, French, Amy J., Thibodeau, Stephen N., Jass, Jeremy, Haile, Robert, Laird, Peter W. (2006). CpG island methylator phenotype

underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nature Genetics 2006 38:7*, *38*(7), 787–793.

Weiser, Martin R. (2018). AJCC 8th Edition: Colorectal Cancer. *Annals of Surgical Oncology*, *25*(6), 1454–1455.

Welsh, Allison W., Harigopal, Malini, Wimberly, Hallie, Prasad, Manju, Rimm, David L. (2013). Quantitative analysis of estrogen receptor expression shows SP1 antibody is more sensitive than 1D5. *Applied Immunohistochemistry and Molecular Morphology*, *21*(2), 139–147.

West, N. P., Dattani, M., McShane, P., Hutchins, G., Grabsch, J., Mueller, W., Treanor, D., Quirke, P., Grabsch, H. (2010). The proportion of tumour cells is an independent predictor for survival in colorectal cancer patients. *British Journal of Cancer 2010 102:10*, *102*(10), 1519–1523.

Wiland, Homer O., Shadrach, Bonnie, Allende, Daniela, Carver, Paula, Goldblum, John R., Liu, Xiuli, Patil, Deepa T., Rybicki, Lisa A., Pai, Rish K. (2014). Morphologic and molecular characterization of traditional serrated adenomas of the distal colon and rectum. *The American Journal of Surgical Pathology*, *38*(9), 1290–1297.

Williams, Bethany, Brettle, David, Aslam, Muhammad, Barrett, Paul, Bryson, Gareth, Cross, Simon, Snead, David, Verrill, Clare, Clarke, Emily, Wright, Alexander, Treanor, Darren. (2020). Guidance for Remote Reporting of Digital Pathology Slides During Periods of Exceptional Service Pressure: An Emergency Response from the UK Royal College of Pathologists. *Journal of Pathology Informatics*, *11*(1).

Winkler, Juliane, Abisoye-Ogunniyan, Abisola, Metcalf, Kevin J., Werb, Zena. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nature Communications 2020 11:1*, *11*(1), 1–19.

Wong, Sunny H., Yu, Jun. (2019). Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nature Reviews Gastroenterology & Hepatology 2019 16:11*, *16*(11), 690–704.

Xu, Hongming, Cha, Yoon Jin, Clemenceau, Jean R., Choi, Jinhwan, Lee, Sung Hak, Kang, Jeonghyun, Hwang, Tae Hyun. (2022). Spatial analysis of tumor-infiltrating lymphocytes in histological sections using deep learning techniques predicts survival in colorectal carcinoma. *The Journal of Pathology: Clinical Research*, *8*(4), 327–339.

Yamane, Letícia, Scapulatempo-Neto, Cristovam, Reis, Rui Manuel, Guimarães, Denise Peixoto. (2014). Serrated pathway in colorectal carcinogenesis. *World Journal of Gastroenterology : WJG*, *20*(10), 2634.

Yamashita, Rikiya, Long, Jin, Longacre, Teri, Peng, Lan, Berry, Gerald, Martin, Brock, Higgins, John, Rubin, Daniel L., Shen, Jeanne. (2021). Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. *The Lancet Oncology*, *22*(1), 132–141.

Yang, Shi, Farraye, Francis A., Mack, Charline, Posnik, Oksana, O'Brien, Michael J. (2004). BRAF and KRAS Mutations in hyperplastic polyps and serrated adenomas of the colorectum: relationship to histology and CpG island methylation status. *The American Journal of Surgical Pathology*, *28*(11), 1452–1459.

Yang, Yuchong, Huang, Xuanzhang, Sun, Jingxu, Gao, Peng, Song, Yongxi, Chen, Xiaowan, Zhao, Junhua, Wang, Zhenning. (2015). Prognostic Value of Perineural Invasion in Colorectal Cancer: A Meta-Analysis. *Journal of Gastrointestinal Surgery*, *19*(6), 1113–1122.

Yoshida, Go J. (2020). Applications of patient-derived tumor xenograft models and tumor organoids. *Journal of Hematology & Oncology 2020 13:1*, *13*(1), 1–16.

Yosinski, Jason, Clune, Jeff, Nguyen, Anh, Fuchs, Thomas, Lipson, Hod. (2015). *Understanding Neural Networks Through Deep Visualization*.

Yun, Sang Oh, Cho, Yong Beom, Lee, Woo Yong, Kim, Hee Cheol, Yun, Seong Hyeon, Park, Yoon Ah, Huh, Jung Wook. (2017). Clinical Significance of Signet-Ring-Cell Colorectal Cancer as a Prognostic Factor. *Annals of Coloproctology*, *33*(6), 232.

Zamcheck, Norman, Doos, Wilhelm G., Prudente, Romido, Lucie, Benjamin B., Gottlieb, Leonard S. (1975). Prognostic factors in colon carcinoma: Correlation of serum carcinoembryonic Antigen Level and Tumor Histopathology. *Human Pathology*, *6*(1), 31–45.

Zeichner, Simon B., Raj, Naveen, Cusnir, Mike, Francavilla, Michael, Hirzel, Alicia. (2012). A De Novo Germline APC Mutation (3927del5) in a Patient with Familial Adenomatous Polyposis: Case Report and Literature Review. *Clinical Medicine Insights. Oncology*, *6*, 315–323.

Zhang, Baoyi, Yao, Kevin, Xu, Min, Wu, Jia, Cheng, Chao. (2021). Deep Learning Predicts EBV Status in Gastric Cancer Based on Spatial Patterns of Lymphocyte Infiltration. *Cancers 2021, Vol. 13, Page 6002*, *13*(23), 6002.

Zhang, Lu, Shay, Jerry W. (2017). Multiple Roles of APC and its Therapeutic Implications in Colorectal Cancer. *JNCI Journal of the National Cancer Institute*, *109*(8).

Zhao, Ke, Li, Zhenhui, Yao, Su, Wang, Yingyi, Wu, Xiaomei, Xu, Zeyan, Wu, Lin, Huang, Yanqi, Liang, Changhong, Liu, Zaiyi. (2020). Artificial intelligence quantified tumour-stroma ratio is an independent predictor for overall survival in resectable colorectal cancer. *EBioMedicine*, *61*, 103054.

Zhou, Naiyun, Fedorov, Andrey, Fennessy, Fiona, Kikinis, Ron, Gao, Yi. (2017). *Large scale digital prostate pathology image analysis combining feature extraction and deep neural network.*

Zhou, Zhi Hua, Ji, Cheng Dong, Xiao, Hua Liang, Zhao, Hai Bin, Cui, You Hong, Bian, Xiu Wu. (2017). Reorganized Collagen in the Tumor Microenvironment of Gastric Cancer and Its Association with Prognosis. *Journal of Cancer*, *8*(8), 1466–1476.