Early prediction of skin toxicities during radiotherapy using optical and infrared imaging

by

Abby Yashayaeva

Submitted in partial fulfilment of the requirements

for the degree of Master of Science

at

Dalhousie University

Halifax, Nova Scotia

August 2022

Table of Contents

List of Tables

List of Figures

Abstract

Approximately 90% of breast cancer radiation therapy patients experience skin toxicities that are difficult to classify and predict ahead of time. A prediction at the early stages of the treatment would provide clinicians with a prompt to intervene. The objectives of this study are to evaluate the correlation between the skin toxicity occurrences and radiomic features extracted from optical and infrared (thermal) images of skin, and to develop a model for predicting a patient's skin response to radiation. Optical and infrared images of breast cancer patients' chest wall were acquired daily during the course of RT, as well as weekly on the three following weeks after treatment. Skin toxicity assessments were conducted weekly until the final visit. The trends of colour and temperature radiomic features of the skin on the treated area were analyzed, reduced, and used in a machine learning model to predict the patients' skin toxicity grade. A set of nine independent colour and temperature features with significant p-values were reduced from a series of 108. The cross-validation accuracy of the machine learning model remained above 80% and AUC above 0.7 when reducing the input data to include only up to a biologically effective dose of 30 Gy (approximately the first third to first half of the treatment dose). The quantitative analysis of radiomic features was shown to be promising for predicting skin toxicities. This study will continue toward the development of a practical and efficient tool that can be used daily in the radiation therapy treatment room.

List of Abbreviations Used

T – Primary Tumor

N – Regional Lymph Nodes

M – Metastasis

DCIS – Ductal Carcinoma in Situ

LINAC – Linear Accelerator

MLC – Multileaf Collimator

NSH – Nova Scotia Health

IMRT – Intensity-Modulated Radiation Therapy

VMAT – Volumetric Modulated Arc Therapy

CT – Computed Tomography

GTV – Gross Tumour Volume

CTV – Clinical Target Volume

PTV – Planning Target Volume

IGRT – Image-Guided Radiotherapy

CTCAE – Common Terminology Criteria for Adverse Events

ROI – Region of Interest

GLCM – Gray-Level Co-occurrence Matrix

GLRLM – Gray-Level Run Length Matrix

GLSZM – Gray-Level Size Zone Matrix

GLDM – Gray-Level Dependence Matrix

SSR – Sum of Squared Ratios

MAD – Median Absolute Deviation

BED – Biologically Effective Dose

SNR – Signal to Noise Ratio

TP – True Positive

TN – True Negative

FP – False Positive

FN – False Negative

ROC – Receiver Operating Characteristic

TPR – True Positive Rate

FPR – False Positive Rate

AUC – Area Under the Curve

SVM – Support Vector Machine

Acknowledgements

Firstly, I would like to thank my supervisor Dr. James Robar for initiating the project and regularly, attentively and thoroughly guiding each stage. When it came to presenting and discussing results and progress made through the course of the project, the weekly meetings led by Dr. Robar were particularly insightful and inspirational for developing, enriching and improving the contexts of the entire study. I would also like to thank Dr. Hannah Dahn, Dr. Michelle Svatos and Dr. Thalat Monajemi for agreeing to be my committee board members and taking the time to attend our meetings and provide me with useful and constructive feedback.

Angela Henry was extremely friendly and helpful with ordering equipment in a timely and efficient manner. We were fortunate to have Ian Conrad design and build crucial components for positioning and holding the devices used during data collection.

I express my kindest gratitude to all of the staff in the radiation room who participated in the data acquisitions and logging. I would particularly like to thank the lead therapists: Shaun Naugle, Karen Sutherland, Brittney Green, Kenny Zhang, Cayleigh Martell for carrying out the patient data collection, as well as the oncologists: Dr. Hannah Dahn and Dr. Rob Rutledge for taking charge of patient recruitment and making decisions on patient eligibility.

Chapter 1   Introduction

1.1   Canadian Demographics of Cancer

Cancer is an undesirable disease caused by mutations in genes that regulate the way cells grow and divide. Cancerous cells grow uncontrollably and may form masses known as tumours which can potentially spread to other organs in a process known as metastasis. Tumours are classified as either benign or malignant based on several characteristics. Benign tumours grow slowly, have a distinct border, and do not invade surrounding tissues. Malignant tumours may grow rapidly, may not have distinct borders, may invade surrounding tissues and may metastasize [1]. Cancer is the leading cause of death in Canada, with an estimated two in five Canadians diagnosed with cancer in their lifetime, and around one in four Canadians is expected to die from cancer [2]. Breast cancer is the most common cancer in Canadian women, contributing to an estimated 25% of cancer cases and 14% of all cancer deaths in 2022. Men will develop less than 0.01% of the total cases of breast cancer [3].

1.2   Treatment of Breast Cancer with External Beam Radiation Therapy

1.2.1   External beam radiotherapy

Radiation therapy is an important component of cancer treatment, used to treat approximately 50% of all cancer patients [4]. It is the standard treatment for most cases

of breast cancer following breast conserving surgery in which the goal is to eliminate any microscopic remains of the tumour with radiation post-surgery [5], and reduce the risk of cancer recurrence, improving overall patient survival. Radiation therapy uses ionizing radiation to damage cancer cells by forming breaks in the DNA that prevent them from further growing and dividing. The absorbed dose is measured in units of gray (Gy), which reflects the amount of energy that ionizing radiation deposits in a material per unit mass. A dose of 1 Gy is equivalent to 1 joule/kilogram.

In external beam radiotherapy, electrons are accelerated to high speeds in a clinical linear accelerator (LINAC) and collide with a heavy metal target, producing high-energy photons. The photons leaving the LINAC are directed towards the tumour in a particle beam arrangement. The LINAC is equipped with a multileaf collimator (MLC) made of series of tungsten leaves, which almost completely attenuate incident x-rays. Each leaf can move independently in different arrangements to block particles from passing, and shape the field of the beam. A picture a LINAC is shown in Figure 1.1 A), with a close-up of the MLC in the LINAC head in B), specifically the one used to treat all patients in our study at the Nova Scotia Health (NSH) Cancer Center.
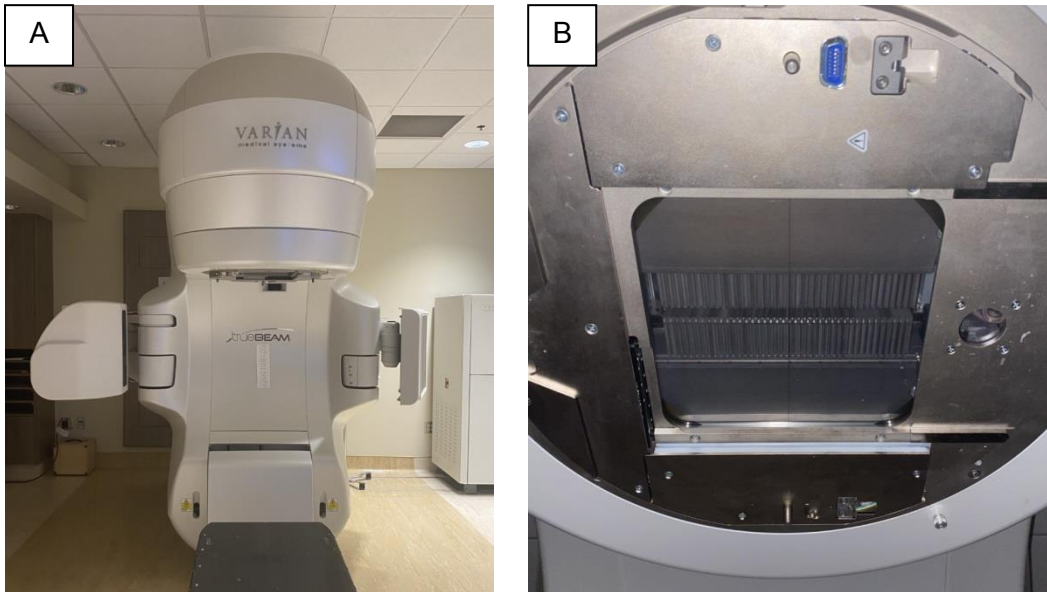
Figure 1.1. A) Medical LINAC and B) MLC on the LINAC at the NSH Cancer Center.

Intensity-modulated radiation therapy (IMRT) is a radiation therapy technique that involves varying the intensity and the shape and orientation of multiple x-ray beams to conform to the shape of a tumor. Volumetric modulated arc therapy (VMAT) is a form of therapy that continuously delivers radiation while the LINAC rotates in a series of arcs around the patient [6] while the MLC shape, the dose rate and the gantry speed are varied in a predefined fashion. The intensity of the beam is shaped by the MLC and modulated at each point by adjusting the dose rate and rotation speed, producing a non-uniform distribution of radiation dose. The combination and overlap of fields from multiple angles of incidence during IMRT and VMAT allows a 3-dimensional dose distribution to be conformed to the tumour volume, while minimizing dose to the surrounding healthy tissues [7]. For breast cancer treatment, tangential field techniques are commonly used in order to provide sparing of main organs-at-risk, including the ipsilateral lung and heart. While early techniques involved fixed, parallel-opposed static fields, more modern

planning techniques may limit the intensity modulation to two opposing tangential fields or two opposing tangential arcs The radiation dose distribution for a left-sided breast cancer patient is shown in Figure 1.2 for a) IMRT and b) VMAT plans, adapted from Karpf et., al [8].
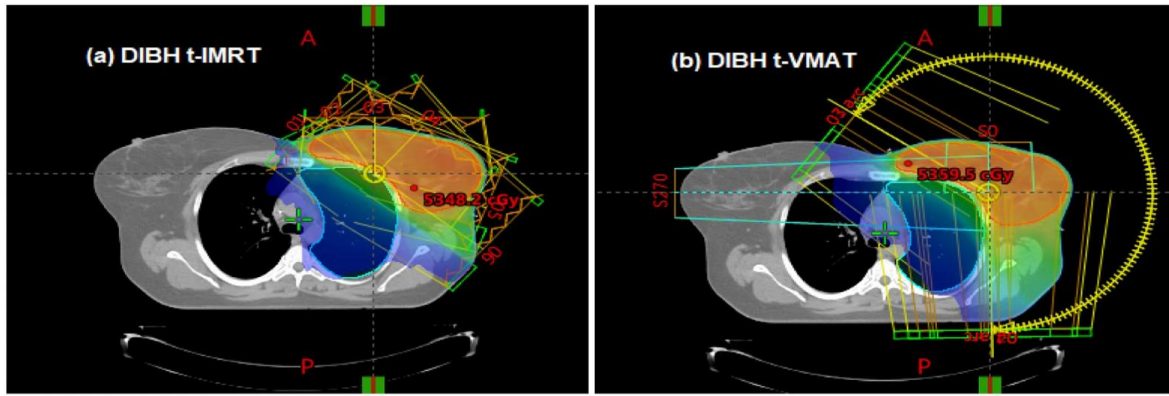
Figure 1.2 a) IMRT and b) VMAT dose distribution for a patient with left-sided breast cancer using a deep inspiration breath hold (DIBH) technique. The t-IMRT plan for this case used 6 equally weighted beams and t-VMAT plan used four arcs, achieving coverage of 95% dose for 95% of the target volume [8].

The total radiation dose is split into a number of treatments called fractions. Delivering smaller doses of radiation over the span of multiple weeks allows for the normal cells to repair between treatments, to reduce side effects. Usually fractions are given five days a week (Monday – Friday) with no treatment over the weekend.

1.2.2   Breast Cancer Staging, Treatment Selection and Fractionation Schemes

Mammography is the most common imaging modality used for screening and detection of breast cancer. Mammography uses low-dose X-rays to examine the breast, producing

an image where cancerous masses appear brighter. After breast cancer has been detected, more detailed evaluations using other modalities, such as breast ultrasound, thermography, magnetic resonance imaging, and positron emission tomography may be performed [16]. A biopsy is then used after the screening to confirm the diagnosis. The diagnosis will help the oncologist decide a treatment approach that provides the best chance of cure and prevent recurrence.

Once breast cancer has been diagnosed, the oncologist will establish the stage of the cancer using imaging techniques and once the primary tumor and lymph nodes have been removed. The stage will help determine the extent of the disease, and the best treatment approach. Staging is determined from information on the size of the primary tumor (T), presence of cancer in regional lymph nodes (N), and the spread of cancer beyond the lymph nodes, known as metastases (M). T is assigned a number between zero and four, with a higher number indicating a larger tumour size that has grown deeper into nearby tissues. N is assigned a number between zero and three, specifying the extent of spread, number and location of nearby lymph nodes containing cancer. Lymph nodes are structures in the lymphatic system that filter damaged cells, like cancer cells, from the lymphatic fluid, and contain immune cells to help fight infections [9]. Regional lymph nodes are the primary sites of lymphatic drainage from the breast, and therefore are considered to be a strong predictor of breast cancer recurrences and survival [10]. These lymph nodes are located under the arm (axillary lymph nodes), near the collarbone (supraclavicular lymph nodes), and near the breastbone (internal mammary lymph nodes) [11]. Figure 1.3 provides a labelled diagram of the regional lymph nodes of the breast.
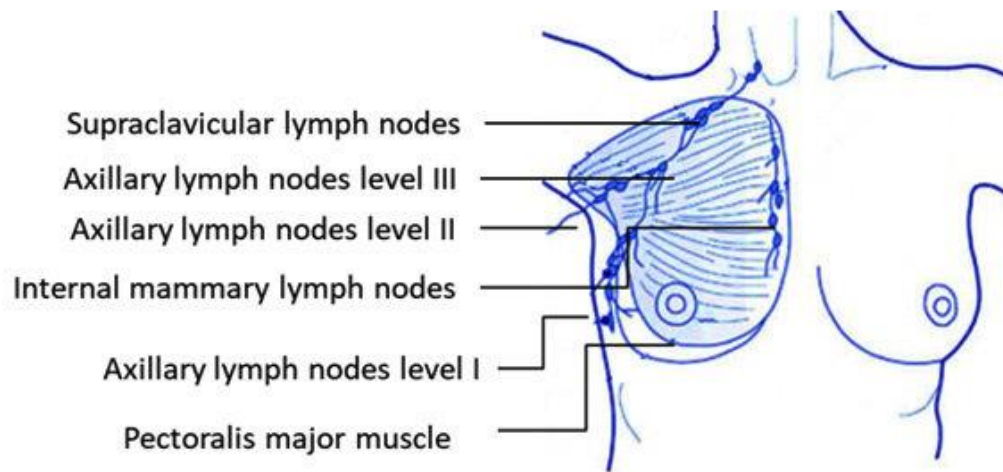
Figure 1.3 Regional Lymph Nodes for Breast [12].

N is the clinical nodal stage based on all imaging and clinical examinations carried out before a surgery to remove the tumor. A pathologic nodal stage (pN) can be determined using microscopic information of tissue removed from surgery or a biopsy. The final component, M, is assigned either zero if the cancer has not spread, or one if it can be detected beyond the regional lymph nodes. The breast cancer-specific TNM assignments are compiled in Table 1.1. [13].

Table 1.1 TNM grading guidelines for breast cancer, adapted from the AJCC 7th edition cancer staging manual [13]. Clinically detected (CD) indicates detected by imaging or clinical examination.

Primary Tumour (T)	Regional Lymph Nodes (N)	Pathologic (pN)	Distant Metastases (M)
TX Primary tumor cannot be assessed.	NX Cancer in regional lymph nodes cannot be assessed.	pNX Regional lymph nodes cannot be assessed	MX Metastasis cannot be measured.
T0 Primary tumor not found.	N0 No cancer in regional lymph nodes.	pN0 No regional lymph node metastasis identified histologically.	M0 No evidence of distant metastases.
Tis Precancerous cells in a location where they were first formed			
T1 Tumor $\leq$ 20 mm in greatest dimension.	N1 Metastases to movable level I, II axillary lymph node(s)	pN1 Metastases in 1–3 axillary lymph nodes and/or in internal mammary nodes with metastases detected by sentinel lymph node biopsy but not CD.	M1 Cancer has spread to other parts of the body.

Primary Tumour (T)	Regional Lymph Nodes (N)	Pathologic (pN)	Distant Metastases (M)
T2 Tumor >20 mm but ≤ 50 mm in greatest dimension.	N2 Metastases in level I, II axillary lymph nodes, or in the internal mammary nodes in the absence of axillary lymph node metastases	pN2 Metastases in 4–9 axillary lymph nodes or in CD internal mammary lymph nodes in the absence of axillary lymph node	
T3 Tumor >50 mm in greatest dimension.	N3 Metastases in the level III axillary lymph node(s); or in the internal mammary lymph node(s) with level I, II axillary lymph node metastases; or in the supraclavicular lymph node(s) with or without axillary or internal mammary lymph node involvement	pN3 Metastases in ≥10 axillary lymph nodes; or in level III axillary lymph nodes; or in CD ipsilateral internal mammary lymph nodes in the presence of positive level I, II axillary lymph nodes; or in ≥3 axillary lymph nodes and in internal mammary lymph nodes with metastases detected by biopsy but not CD; or in ipsilateral supraclavicular lymph nodes	
T4 Tumor of any size with direct extension to the chest wall and/or to the skin			

The AJCC uses different combinations of T, N and M to assign the breast cancer status to nine stages (0, IA, IB, IIA, IIB, IIIA, IIIB, IIIC, and IV) [14]. Stage 0 is a non-invasive precancerous stage where the cancer remains within its original location, and stage IV is an invasive cancer that has spread outside the breast through the blood or lymphatic system to other parts of the body [15]. The TNM combinations for the nine staging assignments are presented in Table 1.2.

Table 1.2 The TNM combination assignments for the nine breast cancer stages, adapted from AJCC 7th edition [13].

Stage	Primary Tumour (T)	Regional Lymph Nodes (N)	Distant Metastases (M)
Stage 0	Tis	N0	M0
Stage IA	T1	N0	M0
Stage IB	T0 T1	N1 N1	M0 M0
Stage IIA	T0 T1 T2	N1 N1 N0	M0 M0 M0
Stage IIB	T2 T3	N1 N0	M0 M0
Stage IIIA	T0 T1 T2 T3 T3	N2 N2 N2 N1 N2	M0 M0 M0 M0 M0
Stage IIIB	T4 T4 T4	N0 N1 N2	M0 M0 M0
Stage IIIC	Any T	N3	M0
Stage IV	Any T	Any N	M1

The selection for suitable breast cancer treatment is based on the stage of the cancer.

Stage 0 Breast Cancer

Approximately 20% of breast cancer cases are non-invasive (stage 0), the most common being Ductal carcinoma in situ (DCIS). DCIS begins in thin tubes in the breast that carry

milk, known as ducts. At this stage, the cancer cells have not spread to nearby tissues and are confined to the ducts [16]. The primary treatment for DCIS is breast-conserving therapy which involves surgery to remove the affected area of the breast and a margin of healthy surrounding tissue (lumpectomy), followed by radiation therapy. A commonly used dose fractionation for whole breast radiation therapy is 40 Gy delivered in 15 daily fractions [17]. Some clinics are increasingly offering whole breast short-course radiation therapy which uses 26 Gy in 5 daily fractions and has similar tumour control as that of standard fractionation [18]. Given that 44% to 86% of local recurrences occur close to the tumor bed [19], a partial breast treatment technique may be offered to patients with small tumours if there are no signs of cancer in the lymph nodes. Partial breast irradiation localizes the radiation specifically to the area from which the tumor was removed, with a margin, minimizing exposure to normal tissue. Candidates for partial breast irradiation may receive a dose of 38.5 Gy in 10 fractions. Surgery to remove the entire breast (mastectomy) is an option for patients if there is a reason to avoid radiotherapy such as receiving prior radiation to the same area, pregnancy, pre-existing conditions [17], as well as the inconvenience of needing to come in to the clinic e.g., if the patient lives far away.

Stage I and II Breast Cancer

Most patients with stage I and II breast cancer are candidates for breast conserving therapy [16]. This includes a lumpectomy as well as a lymph-node biopsy to determine whether cancer has spread beyond a primary tumour, followed by whole breast radiation therapy with a total dose of 40 Gy in 15 daily fractions, or partial breast irradiation for

patients with stage I breast cancer. Patients who have a large tumor relative to the breast size, are pregnant, or have concerns about resulting radiation toxicities might prefer to have a mastectomy and avoid radiation therapy [16, 20].

Stage III and IV Breast Cancer

Locally advanced breast cancer requires several modalities of treatment; chemotherapy, mastectomy and radiation therapy. For radiation therapy, the target volume should include the breast, chest wall and draining lymphatics [16]. A dose of 40 Gy in 15 daily fractions is delivered to these targets, and a boost of 10-12.5 Gy in 5 fractions may be delivered to the area where the primary tumor was located. The purpose of the boost is to destroy any remaining breast cancer cells that could lead to a recurrence, in some high risk patients (ie. age < 50 yrs, multiple risk factors for recurrence) [21]. An extended fractionation schedule of 50 Gy in 25 daily fractions or 50.4 Gy in 28 daily fractions could be recommended for patients with large breast size, or infections/inflammation [17].

1.2.3 Clinical Methodology of Breast Cancer Radiotherapy

Before the radiation treatment begins, the patient will go through a treatment planning process called simulation. During the simulation, the patient will be set up in their treatment position, which in conventional whole breast radiation therapy, is the supine position (lying horizontally with the face and torso facing up) [22]. Tattoos will be marked on the patient's skin to ensure the same position can be reproduced every day for treatment. A breast board, like the one shown in Figure 1.4 is used as an immobilization device to keep the patients arms out of the treatment field and limit movement during treatment.

Figure 1.4 Supine breast board used for immobilization during breast cancer radiation therapy

During the simulation a computed tomography (CT) scan, an imaging procedure that creates a detailed 3D anatomical model of the organs, tissues, bones and blood vessels from a series of x-ray images, is taken of the patient. The dosimetrist then delineates

organs in the CT, such as the breast, the lungs, heart liver, and spinal cord. There are three general tumour volumes; the gross tumour target volume (GTV), the clinical target volume (CTV) and the planning target volume (PTV) [23]. The GTV is the visible tumour mass that can be measured or palpated. Since patients have had their tumors removed, there is no GTV in the planning CT. The CTV is the suspected microscopic extension of cancer around the visible tumour, including the entire ipsilateral breast, and it can also include regional lymph nodes when there is a microscopic spread to lymph nodes. The oncologist delineates the PTV which includes margins accounting for organ movement (internal margin), and uncertainties with patient positioning and the planning and treatment delivery (set-up margin). The margins of the PTV ensure that the desired therapeutic dose is delivered to the CTV [23]. The images from the CT scan, including the delineated target volumes and surrounding organs are used by the medical dosimetrists and physicists to plan the course of treatment and calculate the dose. The plan is developed in the treatment software where the appropriate energy, orientation, shaping, and intensity of the beams are established. A typical planning goal is for 95% of the PTV to receive 100% of the dose, with the total coverage of the PTV receiving at least 95% dose. The dose to the surrounding organs is kept as low as possible, and specific dose limits are usually specified by organ [16, 24].

During each treatment session, the therapists position the patient on the treatment couch, often using the tattoos relative to wall and ceiling lasers, and employing the same patient set up and immobilization as was established in the simulation session. Image-guided radiotherapy (IGRT) involves patient imaging, commonly with an imaging system in the treatment room such as an electronic portal imaging device or cone beam CT (CBCT)

attached to the LINAC, during the treatment to accurately guide the radiation to the target. IGRT helps reveal and reduce errors in the displacement of the target position relative the treatment plan established on the initial CT simulation. In this process, the patient position is corrected as needed, and then re-verified. Once the anatomy and target volume is aligned, the radiation therapists initiate the treatment delivery and dose is delivered to the target [20]. The typical duration of the radiation delivery is only a few minutes, however the treatment session can take 15 to 45 minutes including the patient setup and IGRT.

1.3   Radiation Induced Skin Toxicity

Radiation therapy focuses the radiation dose on the tumour as precisely as possible to avoid destroying normal cells, however, normal tissues along the path of the radiation are still prone to a certain degree of damage and can result in a variety of side effects for the patient. In external beam radiotherapy, radiation unavoidably passes through the patient's skin and may lead to an inflammatory skin response. Radiation induced skin toxicity (or radiation dermatitis) is defined as an inflammatory reaction occurring as a result of exposure to biologically effective levels of ionizing radiation, and is classified on a scale of 1–4 according to the Common Terminology Criteria for Adverse Events (CTCAE) v5.0 [25]. Approximately 90% of radiation therapy breast cancer patients experience at least CTCAE = 1. This grade corresponds to faint erythema or dry desquamation, which appears as a rash and may cause sensations of dryness, burning, itching, and discomfort

to the patient [26]. For over 30% of breast cancer patients, erythema progresses to become more severe and painful moist desquamation (CTCAE = 3) [27, 28], which is characterized by scaling, peeling, bleeding and blistering of the skin and increases the risk of infection. Moist desquamation is strongly associated with long term fibrosis and permanent skin damage, resulting in a decreased Quality of Life [29, 27]. In extreme cases, it can even interrupt the course of therapy, thereby compromising the probability of cure. Table 1.3 summarizes the CTCAE grades of radiation dermatitis [25].

Table 1.3 CTCAE version 6.0, description of radiation dermatitis [25].

Grade 1	Faint erythema or dry desquamation
Grade 2	Moderate to brisk erythema; patchy moist desquamation, mostly confined to skin folds and creases; moderate edema
Grade 3	Moist desquamation in areas other than skin folds and creases; bleeding induced by minor trauma or abrasion
Grade 4	Life-threatening consequences; skin necrosis or ulceration of full thickness dermis; spontaneous bleeding from involved site; skin graft indicated
Grade 5	Death

An example of a patient's skin condition worsening over the course of the treatment is shown in Figure 1.5.
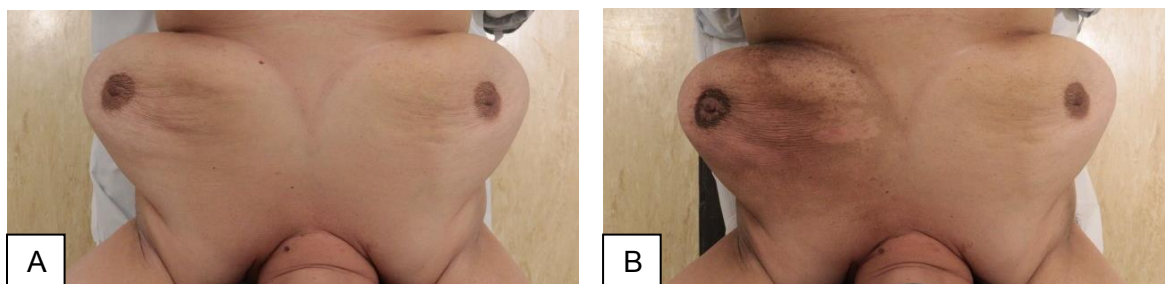
Figure 1.5 Visible change in skin condition between image taken during A) the day of the first treatment session, and B) the day of the last treatment session. The patient's skin toxicity grade assessments during these visits were A) CTCAE = 0 and B) CTCAE = 3.

Moist desquamation occurs more frequently in the weeks after the end of treatment [27, 30]. Figure 1.6 shows the incidences of skin symptoms over time, extracted from Pignol et., al, 2008 [27]. In this study, desquamation peaked during the week after the end of radiation therapy, and the cumulative frequency of erythema and desquamation was around 90% and 30%, respectively. The intensity and rate of skin reaction depends on the thickness and frequency of bolus used (e.g., in Pignol, 2015 [30], over 40% of post-mastectomy radiation therapy patients treated with bolus experienced most desquamation).
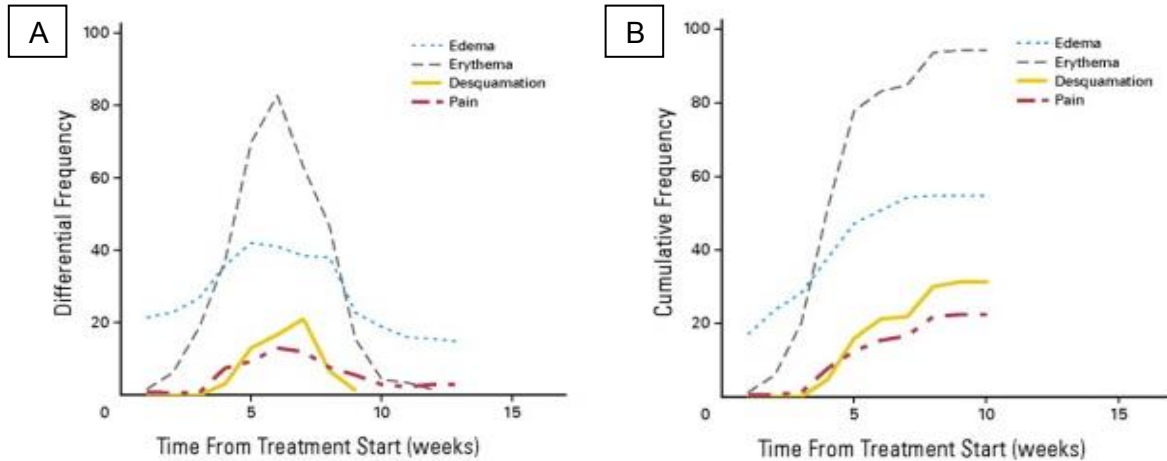
Figure 1.6 A) Differential and B) cumulative incidences of Common Terminology Criteria for Adverse Effects for edema, erythema, desquamation and pain. The data shown begins from one week after the start of radiation therapy [27].

1.3.1 Clinical Difficulties in Pre-emptive Skin Toxicity Mitigation

Each patient experiences a different skin toxicity based on many factors, including treatment parameters such as radiation dose, fractionation, target volume, and patient specific details like age, skin condition, and genetics factors [31]. The dependence of skin reaction on multiple physiological factors and treatment specifics makes it difficult to accurately classify and predict skin toxicity outcomes ahead of time in order to determine the best curative approach, or make adjustments in the treatment plan in extreme situations. In addition, there is currently no method, other than visual examination and subjective diagnosis and prognosis, to quantify and detect early skin reactions that may lead to painful moist desquamation, and eventually permanent side effects. A new automated method for monitoring, staging and ultimately predicting skin response is

needed. If skin toxicity could be predicted early in the course of radiotherapy, a number of options exist to intervene and lower the toxicity for the patient. For example, patients could be informed about their predicted outcomes and advised to begin skin care early on (ie. moisturizer and saline soaks), the clinician could choose to prescribe a steroid cream [32], or adjust or eliminate the use of bolus to lower superficial dose.

1.3.2 Optical and Thermal Imaging in Early Detection of Skin Toxicity

Multiple studies have shown that skin temperature, texture and colour changes are associated with the occurrence of skin toxicity in radiation therapy patients. In a study by Maillot et. al, (2018) [33], temperature rise was recorded for breast cancer patients undergoing treatment, and an association between skin toxicity and temperature increase at the end of radiotherapy was established. Patients who experienced a grade of skin toxicity greater than two had a significantly larger temperature difference between treated and non-treated areas on the skin, on their last treatment visit, compared to patients with a skin toxicity of one or zero. The temperature differences between the set of patients were most likely linked to the inflammatory reaction of the patients' skin [33]. In an earlier study by Templeton et. al (2012) [34], thermal imaging was done to analyze thermal effusivity changes (a measure of temperature or heat change) in irradiated mice. The high grade skin toxicity group of mice showed an earlier and larger increase of relative average effusivity difference than the low grade group [34]. An increase in the average effusivity was present at each fraction for the high grade toxicity group, while the average effusivity varied and the rise was more gradual for the low grade toxicity group. The correlations demonstrated between visual and temperature changes of the skin

and radiation induced skin toxicity suggest that a daily optical and thermal imaging approach over the course of the patient's treatment may provide quantitative biomarkers to predict skin response to therapy during early stages of treatment.

Following the demonstration of thermal metrics for monitoring the occurrence of erythema in previous studies, Saednia et al [35] evaluated the predictive capabilities of thermal metrics using a random forest classifier. An accuracy of 0.87 on independent test data was achieved at the fifth fraction of treatment for predicting skin toxicity at the end of treatment [35]. However, there were several drawbacks in this work. The patients were only imaged at five-fraction intervals (a total of four images per patient), providing limited temporal resolution in imaging data. There could be fluctuations and outliers when only data from a single fraction is used, and it is difficult to assess the error and stability of those results. Imaging at each fraction would allow the analysis of skin condition trends over the course of treatment and reveal any short-term variations. In their study, skin toxicity assessments were performed once at the end of treatment, although it has been established that toxicity continues to evolve in the weeks following the end of treatment [30]. This methodology limited the correlation of their measurement with the eventual toxicities experienced by patients. With regard to imaging methodology, low-resolution thermal images were exclusively used in the algorithm, without incorporating optical imaging. Including high-resolution optical images would provide input for visual signs of erythema and more accurate metrics of the structure of skin abnormalities. Their study also involved a designated examination room, which could be inefficient to incorporate into the daily clinical workflow.

1.4 Optical and Thermal Imaging

1.4.1 Optical and Thermal Image Acquisition

In an optical camera, light rays entering the lens are redirected using mirrors to a single point on a digital camera sensor to create a sharp image. A Digital Single Lens Reflex (DSLR) camera uses a mirror mechanism to reflect light from the camera lens to an optical viewfinder, allowing the photographer to see the image before taking it. When the shutter is released, the mirrors swing out of the light's pathway, and the light rays are incident upon the image sensor to capture the image [36]. Digital cameras use electronic a large number of square cells or pixels, containing light-sensitive sensors. The pixels form a rectangular array, and their number determines the resolution of a given camera and resulting image. Photons hitting a photo sensor create an electrical charge, and the accumulated charge is proportional to amount of light received by the sensor. Each pixel has three types of colour sensors in red, green and blue dyed cells arranged in a mosaic forming a colour filter array. The signals from the colour sensors are interpolated to achieve proper spatial synchronization between the colors. Combining the three color channels allows to recover the full color image [37, 38]. The camera processor converts the image information from the sensor into an appropriate format and writes it into a memory card.

Thermal cameras detect the heat (infrared) energy emitted from an object. Inside of a thermal camera, there are measuring detectors called microbolometers in each pixel of the sensor array that capture infrared radiation. The camera processor takes the signal from the microbolometer and applies a mathematical equation to assign the temperature

in each pixel to an appropriate color, converting it into an electronic image. The resulting matrix of colours is written to memory as a thermal image. The resolution of thermal images from infrared cameras is low in comparison to visible light image because infrared energy has much larger wavelengths than visible light, requiring larger and therefore, fewer sensor elements [39]. Two factors to keep in mind when choosing thermal cameras are and resolution, and thermal sensitivity. Thermal sensitivity is the ability to discriminate between temperature differences, measured in milliKelvins (mK).

1.4.3 Digital Images

An image is a grid (or matrix) of pixels, each having an (x, y) coordinates. In a greyscale image, each pixel is assigned to an intensity ranging from 0 (black) to 255 (white). In a colour image, each pixel is assigned three values corresponding to red, green and blue (RGB). The red, green and blue values of each pixel range from 0 (dark) to 255 (bright) and the combination of these three components can form any colour. For example, a red pixel would have values of R = 255 G = 0 and B = 0. Each colour image can be separated into R, G, and B matrices, known as colour channels [40].

1.5   Radiomics

Medical images (including optical and thermal images in our case) contain characteristics that may not be apparently visible to the naked eye. Uncovering quantitative metrics from the images can provide more information about the tissue of interest. Radiomics is a method that extracts and translates features from medical images using advanced mathematical algorithms to objectively and quantitatively describe observable characteristics [41]. The underlying assumption in oncology is that radiomic features

quantify observable characteristics that are specific to certain conditions. Radiomic features have shown potential in the prediction of response to cancer treatment [42].

The first step in the radiomics algorithm involves identifying a region of interest (ROI) in the images. Next, a large number of radiomic features are extracted from the ROI based on histogram, texture and shape properties, creating a high dimensional feature space. A schematic of this step in shown in Figure 1.7 [42]. Radiomic features can be subdivided into histogram-based (first order), texture-based (higher order), shape-based, and filtered-based features. For the purpose of this study, we will focus on histogram-based and texture-based features.
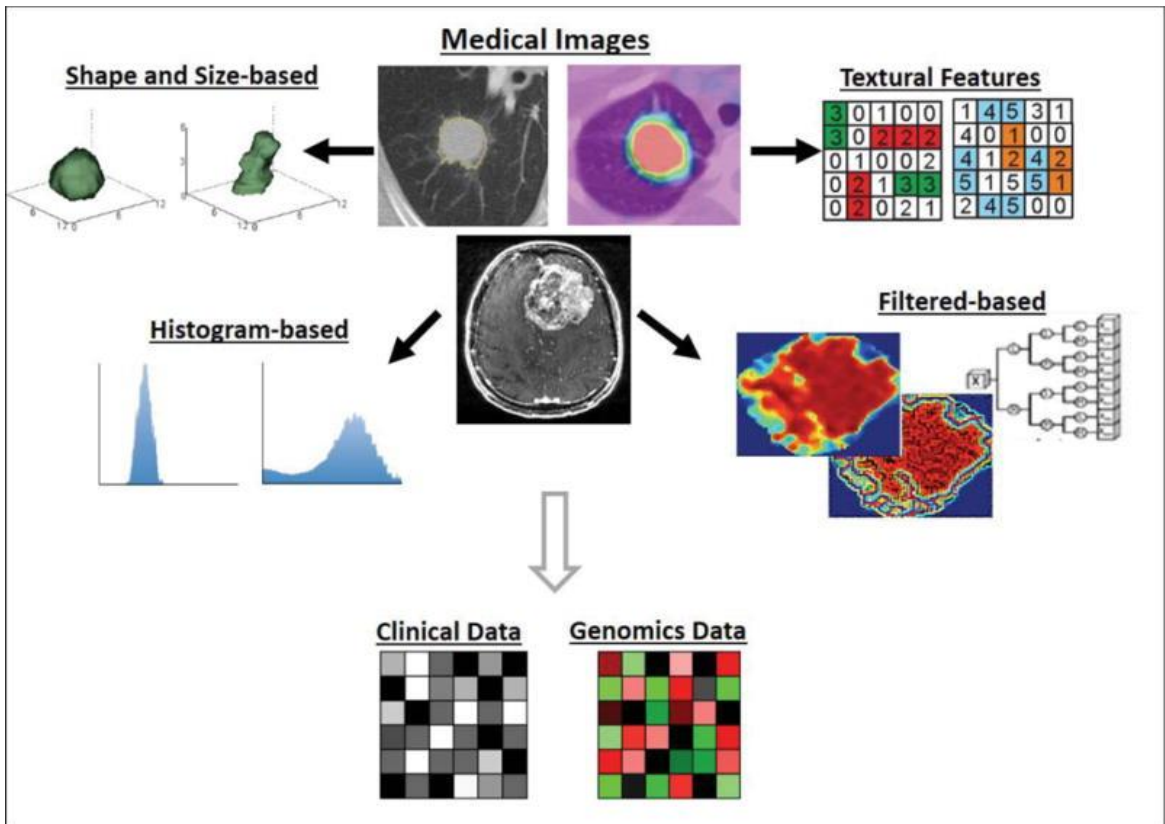
Figure 1.7 Schematic diagram of the radiomic feature extraction, such as shape/size-based, histogram-based, filtered-based, and textural features [42].

Histogram-Based (First Order) Features

The first order statistic features are based on the greyscale histogram that describes the distribution of pixel intensities in an image. These features include gray-level mean, maximum, minimum, variance, and percentiles, and more sophisticated features such as skewness and kurtosis, which describe the shape of the histogram distribution. Histogram-based features are called first order features because they are based on single-pixel analyses [43].

Texture-Based (Higher Order) Features

The first order features describe the distribution of pixel intensities, but do not provide any information about the inter-pixel relationships within the image. The relationship between pixels in an image can be quantified using texture matrices such as the Gray-Level Co-occurrence Matrix (GLCM), Gray-Level Run Length Matrix (GLRLM), Gray-Level Size Zone Matrix (GLSZM), and Gray-Level Dependence Matrix (GLDM). The spatial distribution of intensities within an image can then be described by features extracted from the texture matrices [43].

The GLCM is constructed by considering the relationship between all pixel pairs in the ROI and the frequency of the intensity of each pair. The relationship of two pixels is characterized by the distance and angle between each other [41, 43]. GLCM effectively calculates how often pairs of pixel with specific values and in a specified spatial relationship occur in an image. The frequency of each pixel pair is recorded in the matrix

with positional elements (i,j) corresponding to the first number in the pair (i) and the second number in the pair (j).

GLRLM considers the number of neighbouring pixels (runs) that have the same gray level value (called the gray level run length) in any direction. The positional elements (i,j) in the matrix are the number of adjacent pixels (j) a gray level intensity value (i), appears in the image [41, 43]. GLSZM is similar to GLRLM, but instead of runs, the size of the area of neighbouring pixels with the same grey level value is evaluated [41, 43]. The schematic calculations from the original image matrix into the resulting GLCM, GLRLM and GLSZM are shown in Figure 1.8.
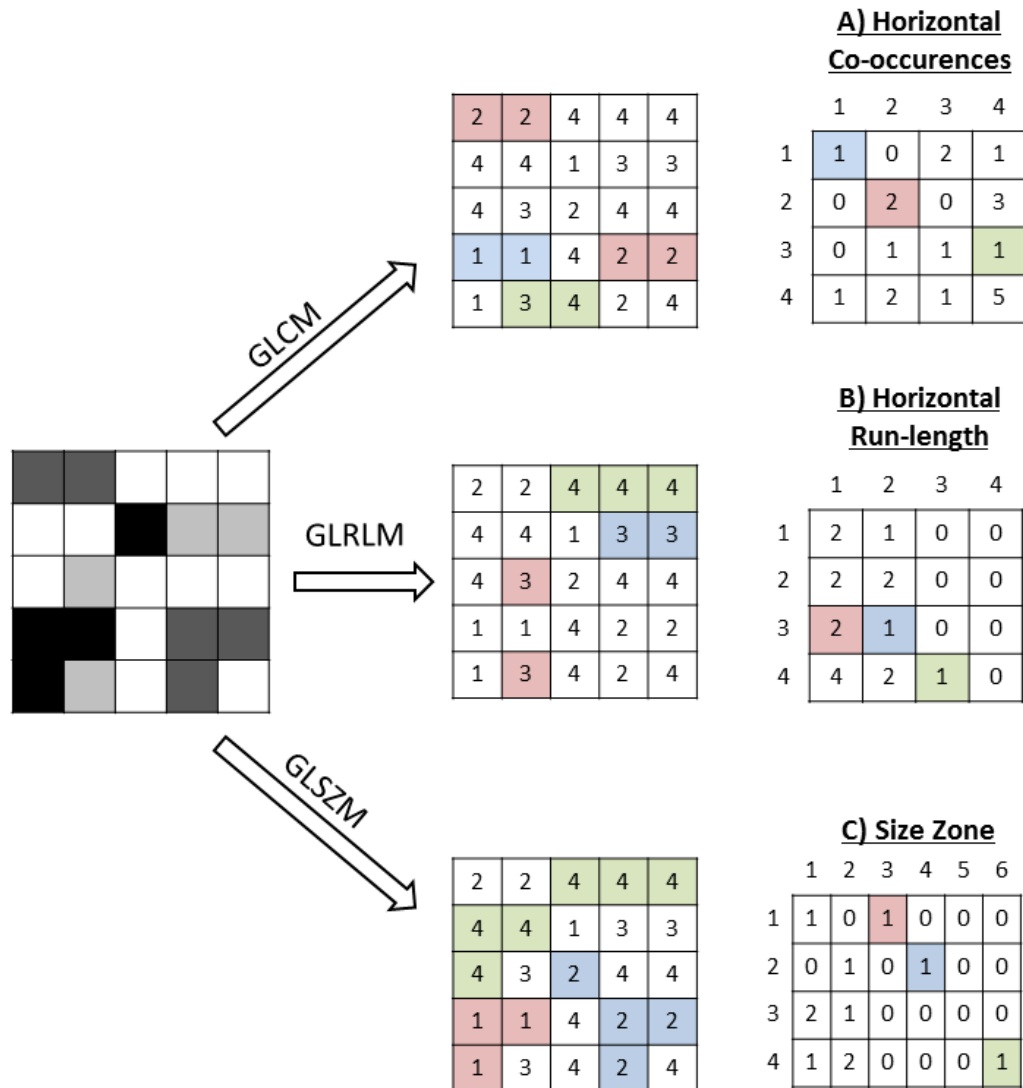
Figure 1.8 Calculation of texture matrices for A) GLCM, B) GLRLM, and C) GLSZM.

GLDM computes the gray level dependencies which is the number of connected pixels that have pixel intensity differences less than a pre-defined value, i.e., a pixel with gray level intensity value j is dependent on its neighbour with a gray level intensity value i if $|i-j| \leq \alpha$ [38, 40]. The (i,j) positional elements describe the number of times a voxel with gray level i with j dependent voxels appears in the image.

An underlying issue in radiomics is a high dimensional feature space. A large number of features with limited sample size would add complexity to any model that uses features as predictors, and cause overfitting. This would negatively impact the predictive power of a model. From all radiomic features extracted, the most informative ones are selected and used to classify the data (e.g., into either positive or negative medical condition) [41]. Reduction of these features avoids the potential problem of data redundancy and overfitting, which would reflect noise in the image more than useful statistics [41].

1.6   Supervised Machine Learning

Machine learning is a method that allows computers to detect patterns and "learn" from data to produce a generalized model that can be used to make predictions on previously unseen, unlabeled data. Supervised machine learning models make predictions based on labeled training data that includes both the input and a desired output [44]. The input data is comprised of a number of features, each representing one dimension of the feature space. In our case, the input data consist of radiomic features extracted from a series of optical and thermal images of patients' skin. The value of a feature data point locates that point in its dimension in the feature space. The values of all features for that data point form a feature vector in the feature space. As an output, each vector in the feature space is assigned to either a numeric value (for a regression algorithm) or a label/class (for a classification algorithm) [44, 45]. A classification algorithm is used in our study since the maximum radiation skin toxicity CTCAE grade detected during the course of the study is

associated to the radiomic data for each patient. In our case, the CTCAE grades are used as labels in the classification.

The goal of classification in machine learning is to categorize subjects, using their numeric input variables, into discrete output variables. Commonly, the input training data are split into smaller training datasets and a validation dataset that is similar to the test dataset to achieve generalizability [44]. The validation dataset is left out of the training data, and is fed into the model to provide an unbiased evaluation. The model sees the validation data set while training and uses it to fine-tune higher level hyperparameters, but never "learns" from it [46]. Lastly, a test dataset is used to evaluate the final models and test its performance on a general population.

Support vector machine (SVM) is a widely used supervised learning algorithm for classification problems. Compared to other algorithms, SVM has a higher speed and performs better on a limited number of samples [47]. The goal of SVM is to find a decision line or hyperplane that separates points into classes in an N-dimensional feature space. The decision line separating the classes can either be linear or non-linear. For example, a linear SVM uses a first-degree polynomial to separate the classes, a quadratic SVM uses a second degree polynomial and a cubic SVM uses a third degree polynomial. The points from both classes that are closest to the line are called support vectors, and the distance between the support vectors and the separation line is called the margin. SVM maximizes the margin to achieve more robust classification decisions on unseen data with higher certainty [48]. Figure 1.9 Illustration of the SVM support vectors, optimal hyperplane (separation line), and maximized margin. Figure 1.9 shows a simple example

of separable data and the optimal hyperplane (separation line) defined by the support vectors.
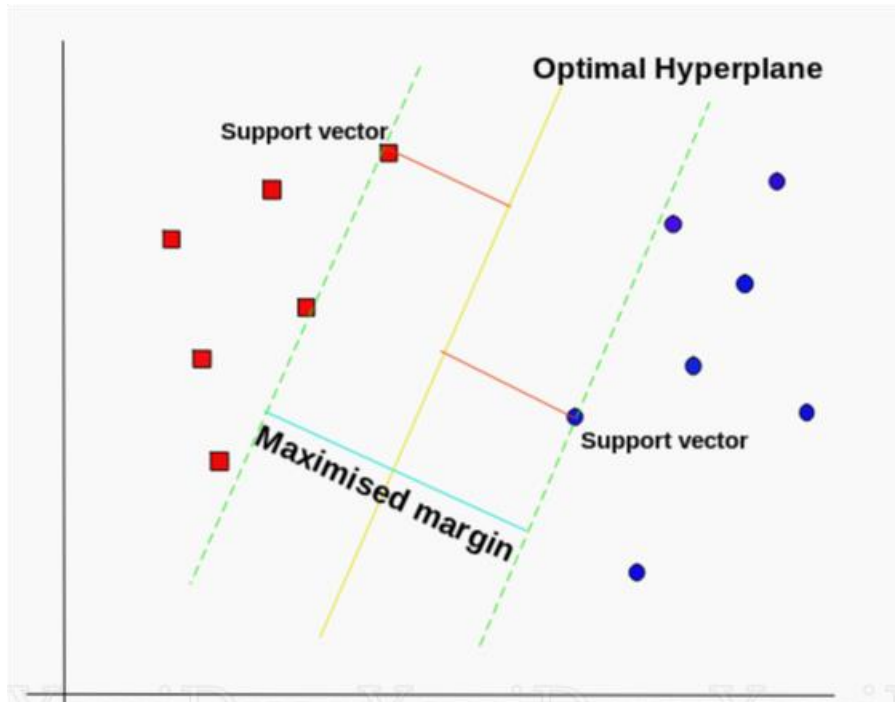
Figure 1.9 Illustration of the SVM support vectors, optimal hyperplane (separation line), and maximized margin [48].

SVM involves optimizing a cost parameter to determine the separation line and width of margin. The cost parameter controls the trade-off between misclassifications and the width of the separation margin between classes [49]. It is a constant multiplied by the number of misclassifications in the machine learning optimization function. The algorithm attempts to minimize the number of misclassifications, while maximizing the margin width. When the cost parameter is large, the margin might be narrow enough to not make any misclassification on the training data, however the separation line will be

bias towards the current data. A smaller cost parameter allows for a degree of violation by the support vector at the margin to achieve a wider margin, and therefore a more generalizable separation line.

Figure 1.10 demonstrates the influence of cost parameter value on the margin and surrounding support vectors.
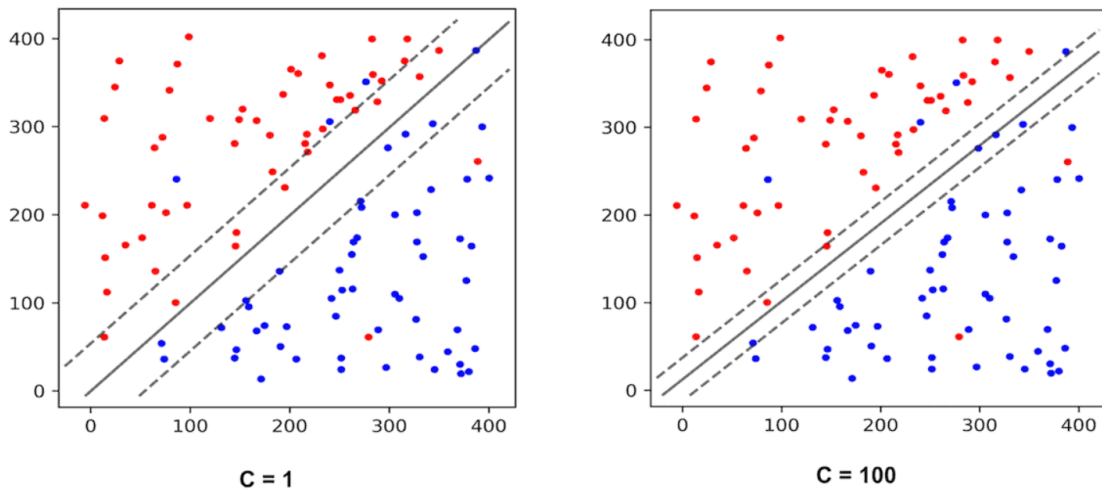
Figure 1.10 SVM produces a wide margin for a small cost parameter (C=1) and narrow margin for a large cost parameter (C=100) [50].

1.7 Study Objectives

We expect trends in spatial, statistical and textural metrics extracted from daily optical images to correlate to changes in images of skin before visual changes are clinically apparent and before symptoms including pain start developing. The combined use of optical and thermal imaging to predict acute skin toxicity has not previously been assessed, nor have image data been acquired frequently enough in this setting to evaluate and utilize trends in the predictive models.

The goal of this study was to define and test a radiomic and machine learning based algorithm to predict maximum patient skin toxicity outcomes based on the assessment of skin color, texture and temperature from daily optical and infrared images taken during the initial few radiotherapy treatment sessions.

Chapter 2 Methods

2.1 Ethical Aspects, Experiment Design, Instrumentation and Data Collection

2.1.1   Ethical approval

This study was registered under ClinicalTrials.gov on April 27th 2020 with ClinicalTrials.gov Identifier: NCT04363892, and approved by the Nova Scotia Health Research Ethics Board on September 25th 2020, assigned to REB File #: 1025729,

2.1.2   Participants

Patients receiving radiation therapy for breast and chest wall cancer of all ages, following breast conserving surgery or mastectomy, were eligible for the study. It was important to include a diverse group of patients in our study to produce a predictive algorithm that could be generalized to be effective for any breast cancer patient undergoing radiotherapy. The patients' ages and treatment-specific details including the total number of fractions, total dose and adjuvant chemotherapy status (yes/no) are summarized in Table 2.1. The maximum CTCAE grades recorded weekly throughout the treatment and 3 follow-up visits were used to classify patients having either minimal symptoms (CTCAE ≤ 1) or moderate to severe symptoms (CTCAE ≥ 2). Only one patient (5%) in our study experienced moist desquamation (CTCAE $= 3$), considerably lower the expected 30% [27, 28], however our sample size may have not been large enough to be representative of moist desquamation incidences.

Table 2.1. Summary of the patient and treatment specific factors for all participants in the study. The p-values are computed from a two-sample student t-Test.

Patient and Treatment Specific Details	CTCAE ≤ 1 (n = 6)	CTCAE ≥ 2 (n = 14)	P-values
Age			0.22
Mean	52	59	
Median	49	60	
Maximum	75	74	
Minimum	39	37	
Standard Deviation	13	11	
Total Fractions			0.51
Mean	16	17	
Median	15	15	
Maximum	20	33	
Minimum	15	15	
Standard Deviation	2	5	
Total Dose			0.30
Mean	4375	4739	
Median	4000	5000	
Maximum	5250	6600	
Minimum	4000	4000	
Standard Deviation	586	739	
Adjuvant Chemotherapy			
Yes	3	1	
No	3	13	

Men were excluded because they rarely develop this form of cancer (less than 1% of all cases). The presence of hair and a different skin complexion would challenge the

development of a texture characterization algorithm, so it was important to include only female patients. Patients with treatment-unrelated skin issues and locally advanced breast cancer were also excluded. While acknowledging that prediction of skin toxicity for patients with abnormal skin conditions and advance breast cancer are equally important, our initial focus was prediction for normal skin conditions first, leaving special cases for follow-up investigations. Any special case would require a dedicated screening of patients and might likely take longer to achieve a sample size sufficient for a proper statistical analysis.

A new list of breast and chest wall cancer patients admitted for radiation therapy was produced weekly, and the primary radiation oncologist for each eligible patient was notified to approach and inform them of the study during their first appointment. If the patient was interested in participating, they were given a form outlining the details of the study to read. If the patient was still interested, at next appointment they signed an informed consent. Each patient participating in the study was given a unique identifier for the anonymization of all data collected.

We had intended to recruit a large enough sample size to assess five factors among patients in a multivariate analysis: skin temperature, skin colour, skin structure, smoking habits, and diabetes. Ideally, having 10 events per factor would comfortably allow us to test and demonstrate its significance, and hence a total of 50 patients experiencing skin toxicities would provide a reliable basis for this factorization. Moist desquamation occurs in approximately one third of the patients, so we initially aimed to recruit 150 patients for assessing the significance of a correlation between each of the listed factors and radiation induced skin toxicity.

In total, over the time available and given limitations during the COVID pandemic, 25 patients consented to the study. Five patients were removed from the study before starting or at different points of their treatment, resulting in a final sample size of 20. Reasons for removal included patients' decision to not attend additional appointments following the treatment course, the oncologist's decision to treat bilaterally (which does not allow imaging of an untreated side) or the decision by the oncologist to treat with a hypofractionated regimen, which would not allow sufficient data acquisition.

2.1.3 Instructional or Intervention Materials

A manual for the data acquisition procedure was prepared for radiotherapists to ensure that consistent conditions were maintained and required steps were properly followed. These instructions included the steps of setting up both cameras on the camera-holding crane over the patient's chest wall, adjusting the room light settings, operating both cameras and adjusting their settings, technique for acquiring images, and storing images on a secure drive.

2.1.4 Instruments

A thermal camera (Forward Looking Infrared FLIR One Pro with 160 x120 infrared thermal resolution and thermal sensitivity of 70 mK) paired and controlled with an iPad, was used to obtain infrared images, and an optical camera (Canon Rebel SL3 DSLR with 24.1 MP resolution) was used to capture optical images of patients' skin. While the thermal camera can also provide registered optical images, the evaluation of the camera performance conducted before the study suggested that the resolution may be too low for the purposes of this study. A plexiglass camera-positioning crane was designed to be secured to the patient couch and to extend over the patient's chest wall 70 cm from the skin surface. For the optical imaging, a series of colour calibration steps were carried out, helping to compensate for any variations in changes in ambient light sources and surface lighting. As part of this procedure, a calibration strip, shown in Figure 2.1, was placed on the patient's skin to appear in a similar region in all images taken throughout the treatment. The colour calibration strips, each containing five colour squares: red, green, blue, black, and white, were printed on cardstock and taped to thin plastic backings to keep them flat to avoid ambient light reflection brightening the strips in images.
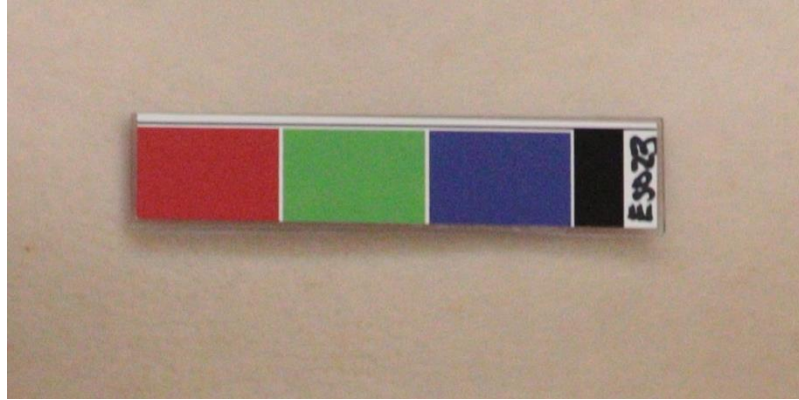
Figure 2.1 Colour calibration strip used for calibrating the intensity of the optical images.

The respective parts of the images were used to calibrate the absolute intensity of each of the three colour channels (red, green, blue) of the optical images. In addition to thermal and optical images skin toxicity assessments were carried out using the CTCAE toxicity grading of skin dermatitis.

Each patient was assigned a secure folder on a protected network drive where their anonymized images and questionnaires were stored. A lockbox attached to the control room wall was used for storing the FLIR/iPad and Canon cameras, while not in use.

2.1.5 Data Collection

Images were collected for patients enrolled in the study daily and throughout the course of their treatment, and weekly for three weeks following the end of treatment. All participants who consented for the study were scheduled to the same treatment room to reduce variability in room brightness and background in the images, as well as to ensure a uniform procedure among staff on the unit. Before the patient arrived for their

appointment, radiation therapists adjusted the size of the aperture on the optical camera to F-stop = 5.6, and sensitivity to ISO = 3600 to ensure consistent brightness and light responsivity. The shutter speed was left on an automatic setting to avoid motion blur. All lights in the treatment room were set to maximum brightness, allowing the optical camera to best detect any visual differences on the skin.

After securing the camera-positioning crane to the treatment unit couch, radiation therapists set up the patient on the couch in in their treatment position. The optical camera was attached to the crane and extended over the patient's chest wall, oriented approximately orthogonal and at a distance of 70 cm from the skin surface to capture both ipsilateral and contralateral sides. The camera position was held fixed with respect to the treatment unit couch for the entire course of the study. The optical camera and iPad-thermal camera pair bolted to the attachments are shown in Figure 2.2, and the camera holding crane with the inserted attachments are shown in Figure 2.3.

Figure 2.2. (A) The FLIR One Pro camera paired to an iPad, and (B) Canon Rebel SL3 DSLR camera used to acquire thermal and optical images, respectively, bolted to plexiglass camera-holding mounts. (C) A close up view of the FLIR One Pro camera.
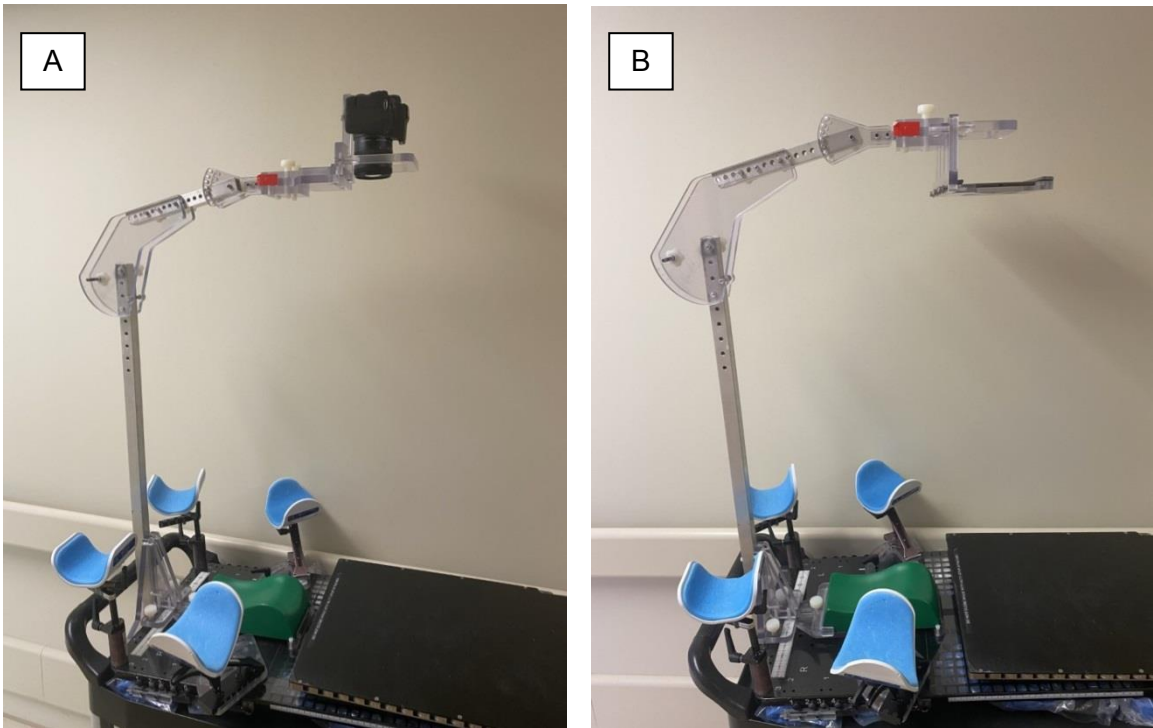
Figure 2.3. The camera-holding crane attached to the end of a breast board with the (A) optical- and (B) thermal-camera attachments inserted.

The colour calibration strip was placed below the patient's chest wall within the field of view of the optical camera to assist with an absolute measurement of the color. Two optical images were acquired (duplication was required to provide both improved statistical robustness and backup in case if the first image was of poor quality) with the self-timer set to three seconds, allowing therapists to step away from the couch to minimize shadowing. Next, the optical camera was replaced with the thermal camera paired to an iPad, and two thermal images of the patient were acquired. A total of 676 optical images and the same number of thermal images were collected for the study. All acquired images were uploaded to a secure folder on a protected network drive and deleted from the cameras immediately after the patient's departure. The thermal camera,

iPad and optical camera were stored and charged within a lock box in the control room of the treatment unit, while not in use.

On the last day of each treatment week and three weeks following the end of treatment, patient questionnaires and CTCAE-based skin toxicity assessments were carried out by radiation therapists.

2.2   Data Processing

As our goal of this work was to design and investigate a practical imaging system that could be implemented at the point of care, we needed to provide means of ensuring consistent colour representation across the optical images, even in the presence of possible variation of ambient light and camera exposure settings. Thus, a method was developed to calibrate all optical images acquired based on sampling of the calibration strip, which was present in all analyzed images.

2.2.1   Automated calibration strip ROI identification

An algorithm was developed for the automated identification of regions of interest (ROIs) for each square in the calibration strips. The sum of squared ratios (SSRs), as defined in equations *(1)-(3)*, were computed for each pixel in the selected area.

$$SSR_{red} = \left(\frac{C_{red}}{C_{green}}\right)^2 + \left(\frac{C_{red}}{C_{blue}}\right)^2 \qquad (1)$$

$$SSR_{green} = \left(\frac{C_{green}}{C_{red}}\right)^2 + \left(\frac{C_{green}}{C_{blue}}\right)^2 \qquad (2)$$

$$SSR_{blue} = \left(\frac{C_{blue}}{C_{red}}\right)^2 + \left(\frac{C_{blue}}{C_{green}}\right)^2 \qquad (3)$$

Where SSR_X is the sum of square ratios for the colour channel of interest X, and C_Y is the spatially dependent pixel value of the Y colour channel.

The SSRs represent the contribution of each colour channel to the total RGB value, relative to the two other colour channels. The colours on the strip were chosen to be red, green and blue were chosen for simple ROI detection using SSRs. For example, a red pixel will have a high red channel value, but low green and blue channel values, so SSR_{red} will be high for that pixel. However, a white pixel would have high red, green and blue channel values, so SSR_{red} of the white pixel will be lower than that of the red pixel.

Next, n pixels with maximum SSR values for each colour channel were selected to cover a threshold of 0.04% of the area with the coordinates $(x_1, y_1), (x_2, y_2), \dots (x_n, y_n)$. The maximum SSR coordinates were used to determine a strength field, defined in equation (4).

$$S(x,y) = \sum_{i}^{n} e^{-\left(\sqrt{(x-x_i)^2+(y-y_i)^2}\right)} \qquad (4)$$

Where $S(x,y)$ is the strength field of the pixel positioned at (x,y) in the image grid, and $\{x_i, y_i\}_i^n$ is the set of coordinates with maximum SSR values.

$S(x,y)$ is greater for pixels located closest to a cluster of maximum SSR points. The points outlined by the contour line of $S(x,y) = 500$ within each ROI were chosen to be used for the colour calibration. The $S(x,y) = 300 \; and \; 500$ contour lines are shown in Figure 2.4 for the A) red, B) green, and C) blue colour channels. The pixels with the lowest and highest combined RGB values were used to capture the darkest and lightest points from the black and white ROIs, respectively.
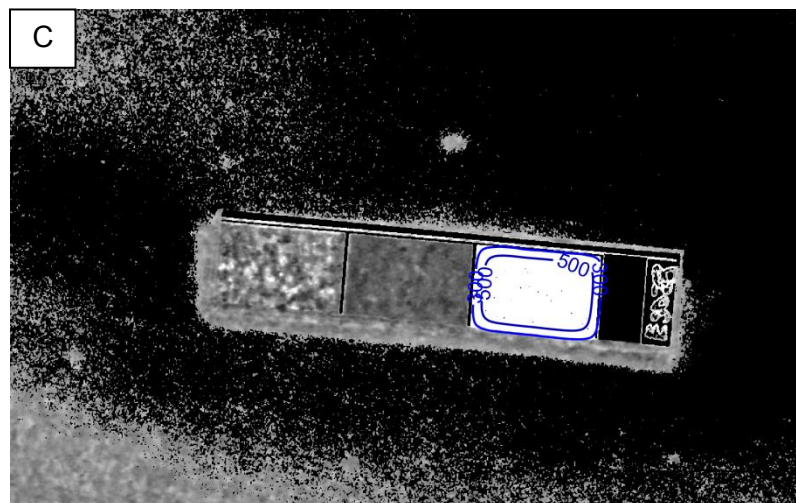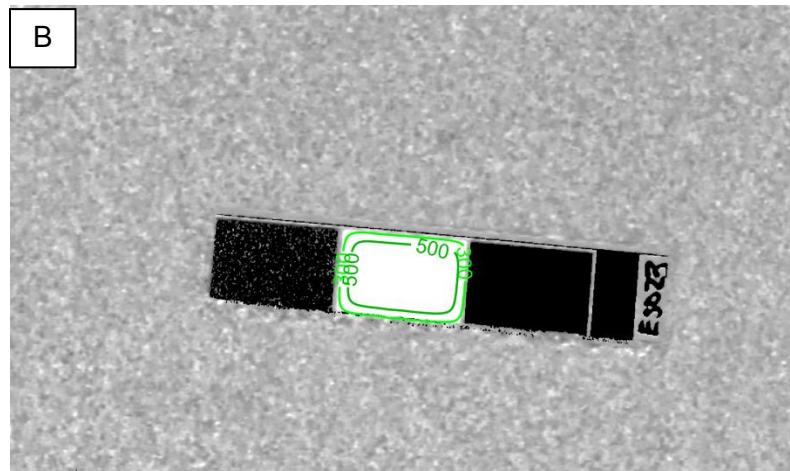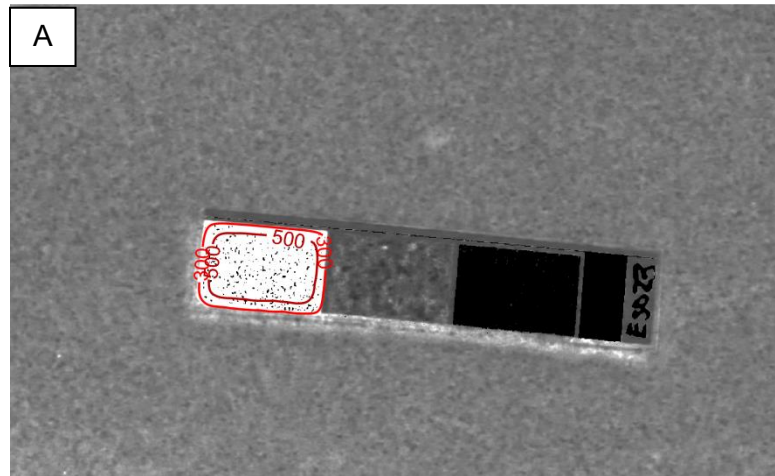
Figure 2.4 The contour lines of the strength field constructed from the SSRs of the A) red, B) green and C) blue colour channel, shown on the colour strip.

2.2.2 Colour calibration from selected ROIs on the Calibration Strip

The median pixel value of the three colour channels within each of the five regions on the strip was taken, resulting in 15 calibration values per image (XY_i), where X indicates the colour of the ROI, Y indicates the colour channel of interest, and i = image index. Figure 2.5 is a diagram of the calibration strip showing the 15 calibration values where the colour abbreviations replacing X and Y are R = red, G = green, B = blue, K = black, W = white.

RR_i	GR_i	BR_i	KR_i	WR_i
RG_i	GG_i	BG_i	KG_i	WG_i
RB_i	GB_i	BB_i	KB_i	WB_i

Figure 2.5. Diagram depicting the calibration values extracted from each ROI in the calibration strip.

The calibration values were averaged across all optical images to get fifteen reference calibration values for the whole image set, defined in equation (5).

$$XY_{ref} = \frac{\sum_i XY_i}{N} \qquad (5)$$

Where N is the number of images in the set. The correction multipliers $(C(XY_i))$ were computed as the ratios of calibration values to the corresponding reference values,

defined by equation (6). Multiplying the RGB pixel values by this factor adjusts colour intensity to match that of the reference colour intensity.

$$C(XY_i) = \frac{XY_i}{XY_{ref}} \qquad (6)$$

Using the assumption that the correction multipliers should be the same among all ROIs and colour channels, a median absolute deviation rejection criterion was applied to remove any outliers from the 15 correction multipliers. The median absolute deviation (MAD) for each image (i) is defined by equation (7) where $\widetilde{C(XY_i)}$ is the median of the 15 correction multipliers.

$$MAD_i = Median\left(\left|C(XY_i) - \widetilde{C(XY_i)}\right|\right) \qquad (7)$$

Any correction multipliers with absolute deviations greater than $4* MAD_i$ were rejected. The remaining correction multipliers were averaged to get one single correction multiplier for each individual image, and calibrate it accordingly. Figure 2.6 shows the correction multipliers values from each ROI, with the colour of the points representing the corresponding colour channel (red, green, blue), and the rejected points marked with asterisks. This is one of the few images that were taken before the calibration strips were glued to the plastic backings, so the red square on the calibration strip displayed visible reflection from the room lights, and therefore gave inaccurate correction multipliers.
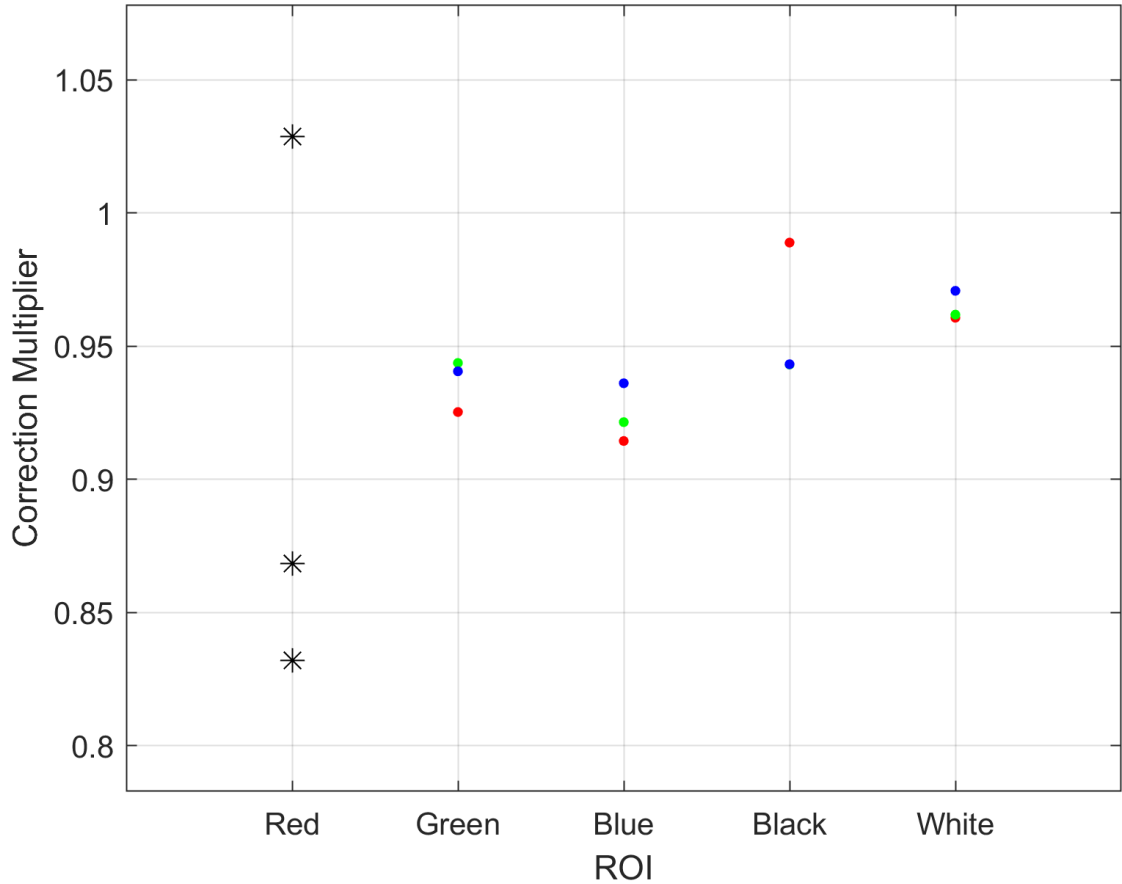
Figure 2.6. Correction multipliers derived from each colour box ROI on the calibration strip in a single image. The red, green and blue points represent the respective colour channels. A median absolute deviation of 4*MAD was used to reject outliers, shown in asterisks (in this case, the correction multiplier from the red ROI were all rejected).

2.2.3 Thermal image calibration

Metadata including the pixel value to temperature conversion coefficients were extracted from the FLIR jpeg files and used to calibrate the arbitrary pixel values of each infrared image into the corresponding absolute temperature values. An accuracy test for the thermal camera was carried out using three wooden blocks and a small piece of carpet (materials with low thermal conductivity) on a flat surface. The relative temperature differences between each block and the piece of carpet were measured five times in stable

room conditions and the standard deviation of the measurements were computed. This test was repeated by increasing the distance from the camera to the surface of the blocks from 19 cm to 70 cm. The relative temperature differences and corresponding standard deviation variation across the tested distances are presented in Figure 2.7.
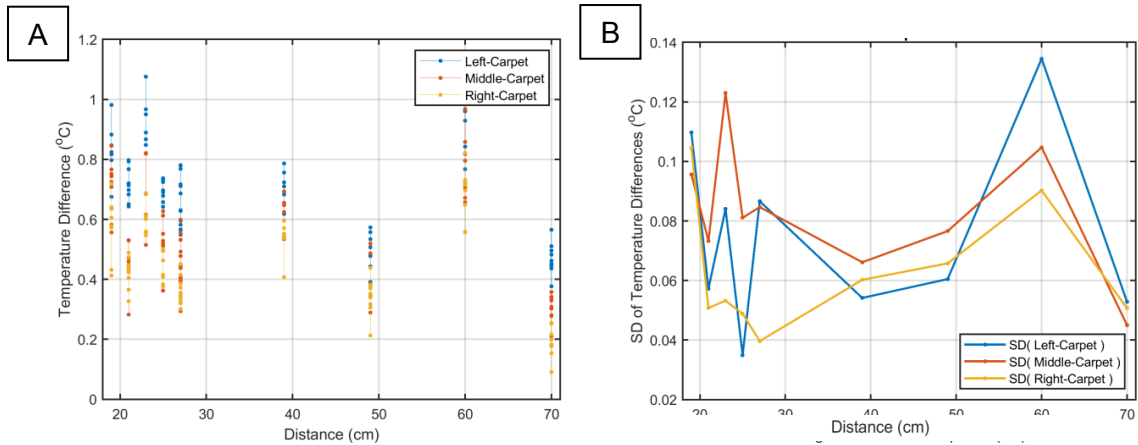
Figure 2.7 A) Relative temperature differences between three wooden blocks and carpet measured five times when increasing the distance between the thermal camera and blocks. The standard deviation of the five back to back measurements are plotted in B) with increasing distance.

The temperature differences between the blocks of wood and carpet within a single image should be close to constant at stable room conditions. The standard deviation of the relative temperature differences remained 0.08°C ± 0.06°C across all distances. Saednia et al reported an increase in mean skin temperature of 0.45°C ± 0.202°C for patients with a skin toxicity grade of CTCAE ≥ 2. The precision of the thermal camera used in our study allows for the detection of the expected relative thermal changes.

2.2.4 Image Registration

To evaluate the optical and thermal skin changes in the same region of the patient's chest at each point throughout the treatment, the images were aligned through a procedure called rigid image registration. This algorithm computes a mutual information metric from a joint probability distribution of two images, which is a function that describes the relationship between the intensities of corresponding pixels in two images. Image registration optimizes the mutual information metric by applying a series of geometric transformations to the image determine one that matches its intensity patterns or features with those of a reference image. The number of samples used to compute the probability density function and number of bins used to compute the mutual information metric were set to 500 and 50, respectively [44]. In our case, an affine transformation was used, allowing translation, rotation, scaling, and shearing of the image. For each patient, an optical and thermal region-of-interest around the patient's entire chest was defined for co-registering the images. One optical image in the image set was chosen as a reference to which all other images were rigidly co-registered to. The reference image was chosen to be at the halfway point of the treatment course to take into account the visual changes of the treated breast at later stages of treatment. The same co-registration procedure was applied to the thermal images. The transformation parameters were then used to align the patient images for each image type individually. Figure 2.8 shows the image transformation before and after affine image registration for both optical and thermal images of a patient taken during two separate visits.
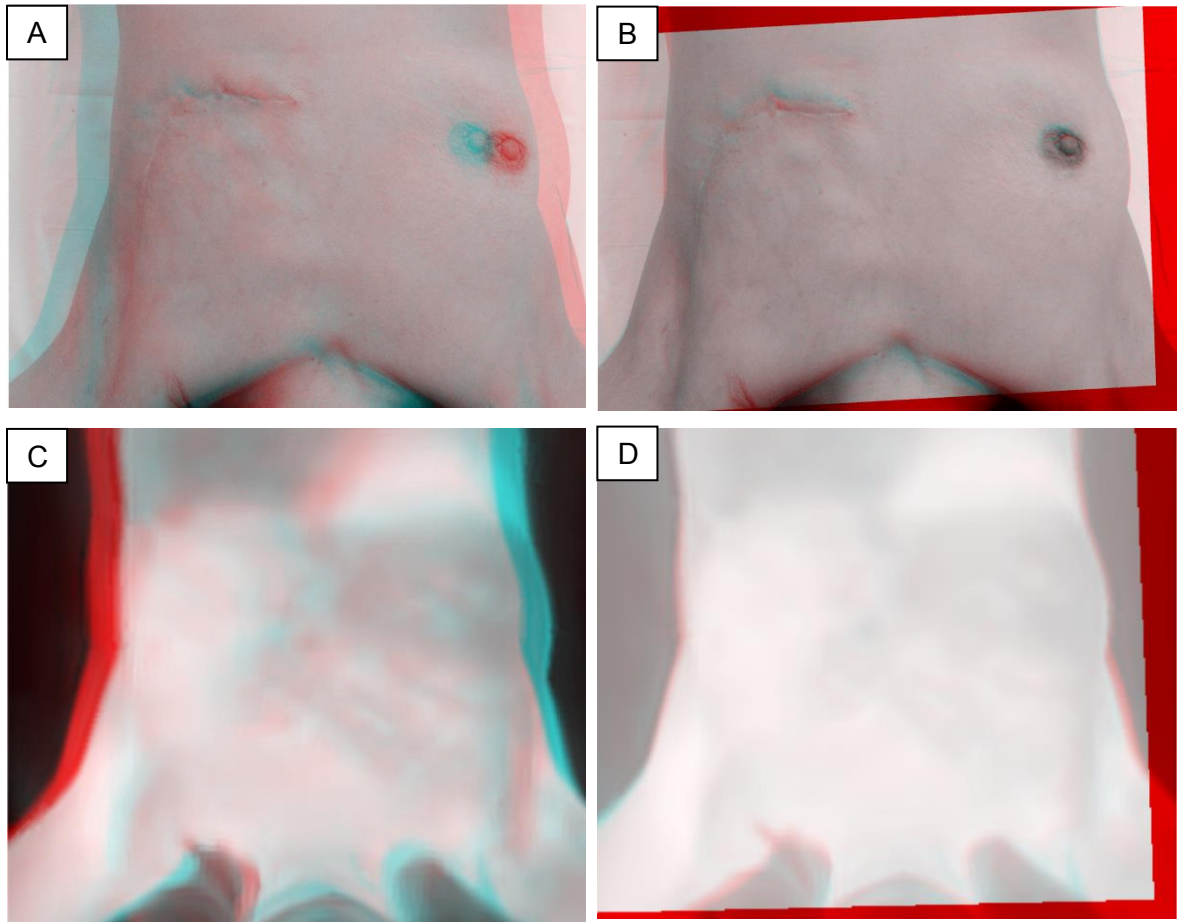
Figure 2.8 A visible shift in the patient's position in the overlaid A) optical and C) thermal images taken during two different visits. The corresponding images are aligned after an affine transformation is applied to the B) optical and D) thermal images.

2.3   Data analysis

2.3.1   Radiomic Feature Extraction

An ROI on the ipsilateral (treated) side of the patient's chest was defined corresponding to the area of skin in the direct field of the radiation beam, using the combined beam projections in the treatment planning system software as a reference. This ROI was used across the entire co-registered optical image set, and the same ROI identification was

repeated for the co-registered thermal image set. A control ROI was outlined on the contralateral (untreated) side, inferior to the patient's clavicle where the colour and temperature appeared the most uniform for each image-type set, and used across the two co-registered image sets. Figure 2.9 is an example of the optical and thermal images for the first patient in the study, with the ipsilateral ROI outlined on the left side, and contralateral ROI outlined on the right side.
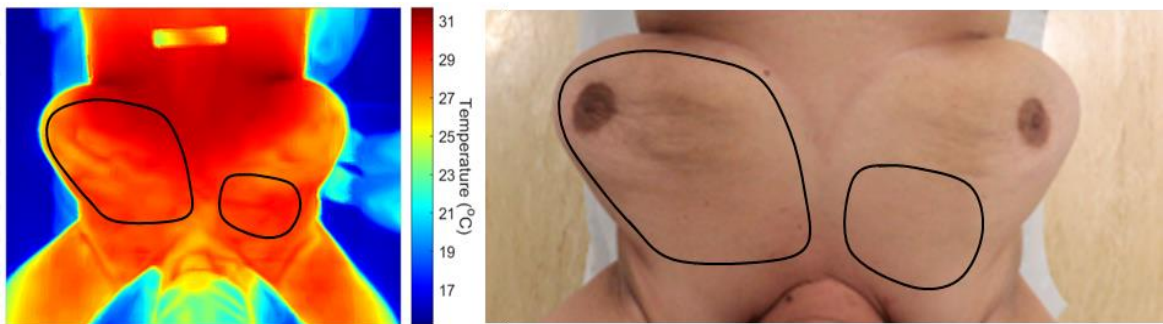
Figure 2.9. The thermal (left) and optical (right) images of a patient's chest with the ipsilateral and contralateral ROI's outlined on each side.

The Python Pyradiomics open-source toolbox was used to extract a series of 108 radiomic features from the ipsilateral and contralateral ROIs on the patients' skin for each image type, for both optical and thermal images. Average values were computed between the duplicate image for all features. Possible trends with biologically effective dose (BED) of these features were examined. BED for each fraction (equation (8)) was computed to allow for comparison between patients with different fractionation schedules. It is defined as a measure of the biological dose delivered by a particular combination of dose per fraction and total dose to a particular tissue characterized by a specific α/β ratio [51].

$$BED = nd\left(1 + \frac{d}{\alpha/\beta}\right) \qquad\qquad (8)$$

In equation (8), n is the number of fractions, d is the dose per fraction, and α/β is the ratio determining the sensitivity of a particular cell type to radiation. For skin erythema and desquamation, $\alpha/\beta = 10.8$ Gy, determined by averaging the α/β values from different studies stated in a review article by Emami et al. [52].

All features were analyzed as function of BED, and their signal-to-noise-ratios (SNR) with respect to linear fit, were computed. In Figure 2.10 the mean gray level feature differences between the two ROIs is shown for a single patient as an example. The differences of the mean ROI values for the optical images show persistent trends as the skin conditions deteriorate when BED increases in Figure 2.10 A). The intensity with respect to the contralateral side declines, which indicates that the colour of the skin is darkening. Likewise, from the thermal images, the differences of the mean ROI values increase as shown in Figure 2.10 B), indicating a rise in temperature of the skin with increasing BED. The high level of variation between fractions (noise) in the thermal data could be due to challenges in executing perfect alignment in thermal image registration. Thermal images are not as detailed as detailed as optical images, and therefore there is more uncertainty when registering all images in the set to a single reference image.
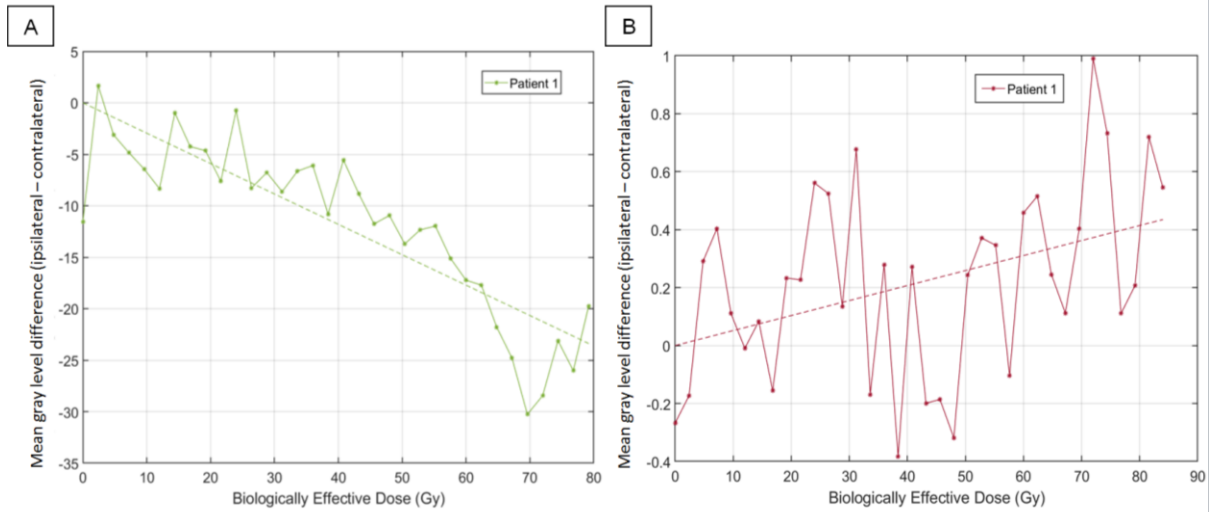
Figure 2.10. The mean gray level ROI differences from (A) optical and (B) thermal with increasing BED corresponding to treatment fractions for the first patient in the study. The zero-intercept is used as a reference point.

Figure 2.10 demonstrates one approach to the analysis of skin characterization. It was based on the assessments of the trends and strength of linear signal for each individual patient. Another approach we followed was taking an ensemble of all patients for a specific feature with the zero-intercept used as the common reference point, and analyzing the probability distribution function for each dose value. This gave an idea of the overall skin response trend of the group, and allowed us to preliminarily group different features by visual comparison of their uniqueness and redundancy, helping to estimate reasonable thresholds for feature reduction. An example of this approach is shown in

Figure 2.11. In each dose-feature value interval, the normalized probability density was computed as the number of cases in the bin divided by the sum of all cases corresponding to the same dose interval, divided by the feature interval size. Because we used the

normalized feature value ranging from zero to one, the feature bin size is equal to one

divided by the total number of intervals. Therefore, the normalized probability density

equals the number of cases in the dose-feature value interval divided by the sum of cases

of that dose interval and multiplied by the number of feature intervals. The normalized

feature range was reassigned to the original range for the probability distribution graphs.

Using the normalized probability function allows comparison between different features.
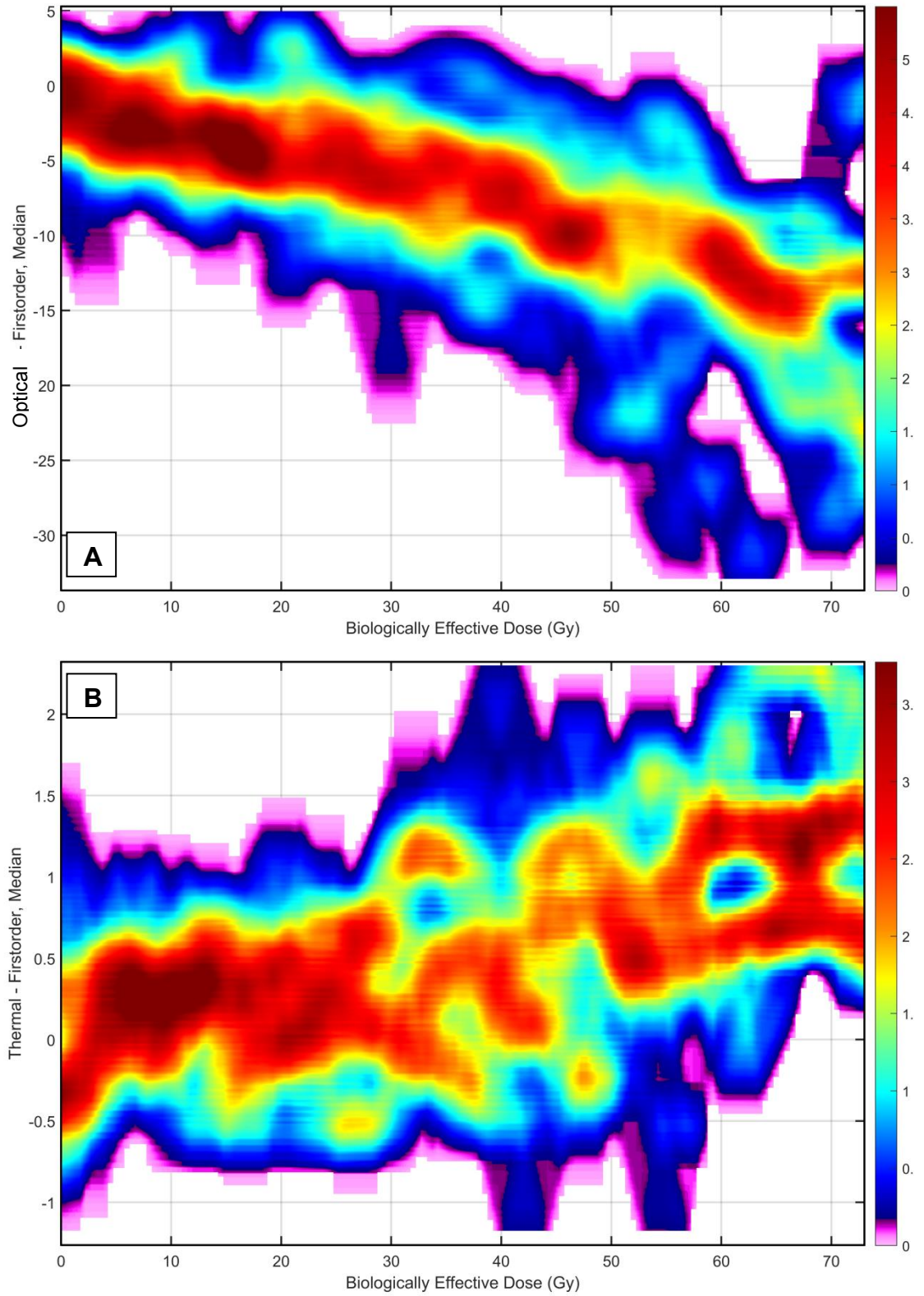
Figure 2.11 Normalized (zero to one range of corresponding feature value) probability distribution function of the median radiomic A) optical and B) thermal features.

The higher probability density shifts to lower median values in the optical images with increasing dose, indicating skin darkening over the course of treatment, On the contrary to that, the thermal median increases with dose, suggesting an increase of skin temperature of the treated area. At a BED of 50 Gy, the distribution becomes bimodal (the peaks in distribution splits), which is apparent in both the optical and thermal median. This, as we later show, corresponds to the separation between high and low skin toxicity grade cases. As mentioned, this method allows to identify the characteristic behaviour of certain features. Figure 2.12 shows two more examples of a systematic change of probability density with dose, but somewhat different than the median feature.
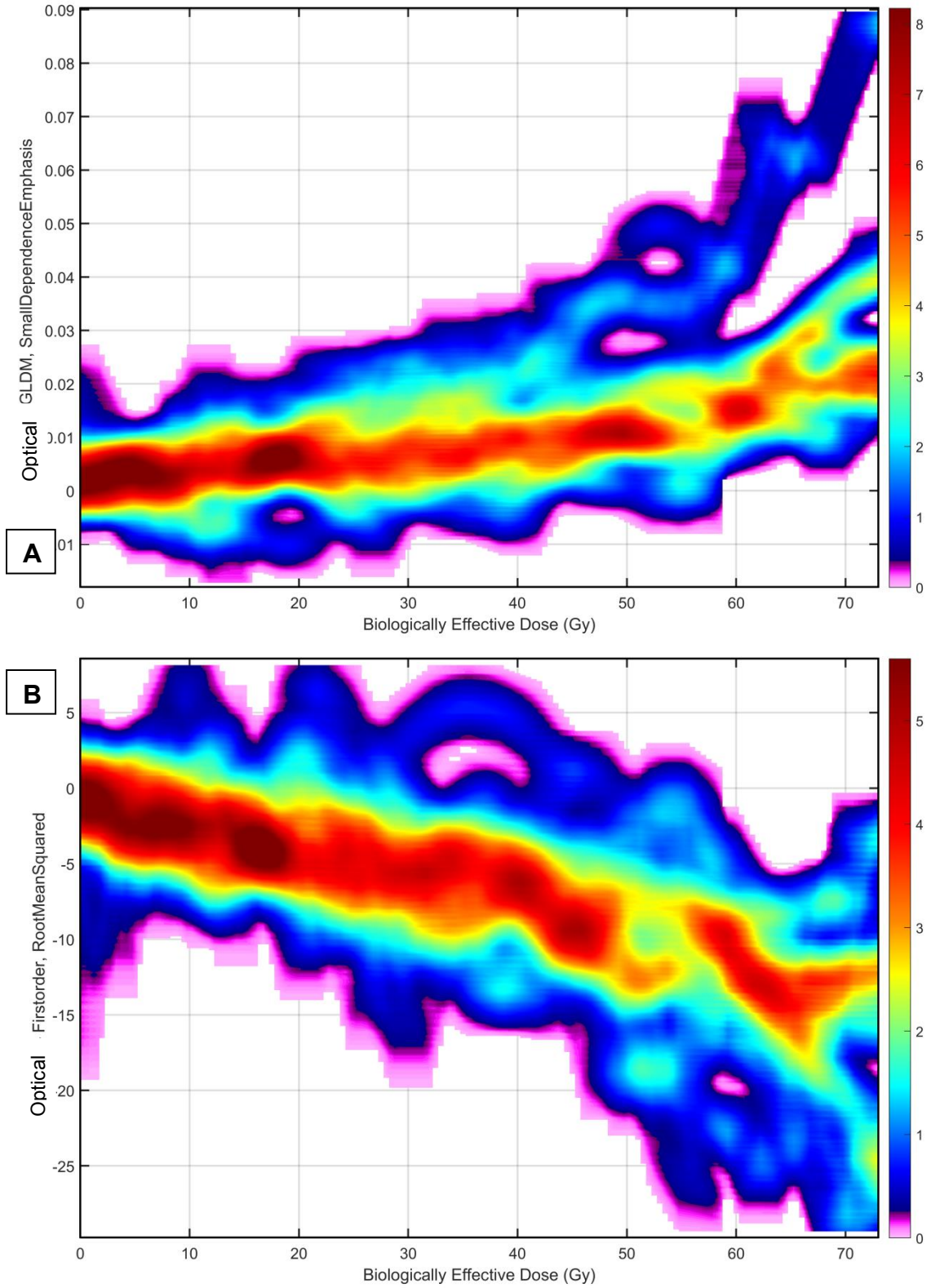
Figure 2.12 Normalized probability distribution function of optical A) first order root mean squared and B) GLDM small dependence emphasis feature.

4.3.2 Feature Quality Evaluation

SNR was defined in equation (9) as the percent of variance removed after subtracting a linear fit (y_{fitted}) from the data points (y). This quantity determined how noisy the data was relative to the trend.

$$\textbf{SNR} = \textbf{100\%} - \frac{\textbf{var}(\textbf{y} - \textbf{y}_{\textbf{fitted}})}{\textbf{var}(\textbf{y})} * \textbf{100\%} \qquad (9)$$

Variance is defined as

$$\textbf{var}(\textbf{x}) = \frac{\textbf{1}}{\textbf{N}} \sum_{i=1}^{N} x_i - \mu \qquad (10)$$

Where N is the number of data points, and μ is the average of all x_i data points.

It was noted that in almost all cases, the SNR of the (ipsilateral − contralateral) ROI feature differences were larger than those of the treated ROI alone. However, the higher order thermal features were an exception to this, demonstrating smaller SNR for the ipsilateral - contralateral ROI feature differences than for the treated ROI features. This may be explained by the reliance of higher order features on high spatial resolution in images. Specifically, the error of the difference of higher order feature values computed for any two nonoverlapping ROIs in a low-resolution thermal image can be expressed as the sum of the random errors of the two feature values. These random errors include instrumental noise, effects of coarse resolution and sensitivity. The cumulative random error of the difference seemingly outweighed the signal of higher order features that is

essentially removed when the difference is taken between the (ipsilateral – contralateral) ROI values. The combined noise when taking differences between higher order features from two different ROIs in the low resolution thermal images outweighed the stabilization from the control ROI. Therefore, first order optical and thermal feature differences and higher order optical feature differences were used in the analysis, while higher order thermal features from the treated ROIs were used directly without performing subtraction of the control/contralateral ROI values.

2.3.3 Feature Dimension Reduction

Some radiomic features may not contribute useful information to the model and are frequently correlated with each other, indicating data redundancy. To avoid overfitting the model and increase classification accuracy, noisy features should be removed from the feature set and correlated features may be grouped and replaced with a representative feature. All shape features were first removed from the set of 108 features, resulting in a set of 83 histogram and textural features. Noisy features were then eliminated by thresholding the average SNR across patients to keep only the features with average SNR equal or greater than = 30% and 15% for optical and thermal images, respectively. This reduced the 83 features from both image types to 28 optical features and 21 thermal features. To help identify and further eliminate linearly dependent variables, the correlation coefficients between each of the feature value series were then computed for both SNR-based reduced image type sets individually. The absolute correlation coefficients are presented in Figure 2.13 (A) and (B) in the form of correlation-matrices of the SNR-based reduced optical and thermal features, respectively, for one patient. The features with higher correlation coefficient values are considered as closely linearly

dependent, and also exhibit similar row and column patterns within the matrix. Both matrices in Figure 2.13 show groups of dependent (highly correlated) features.
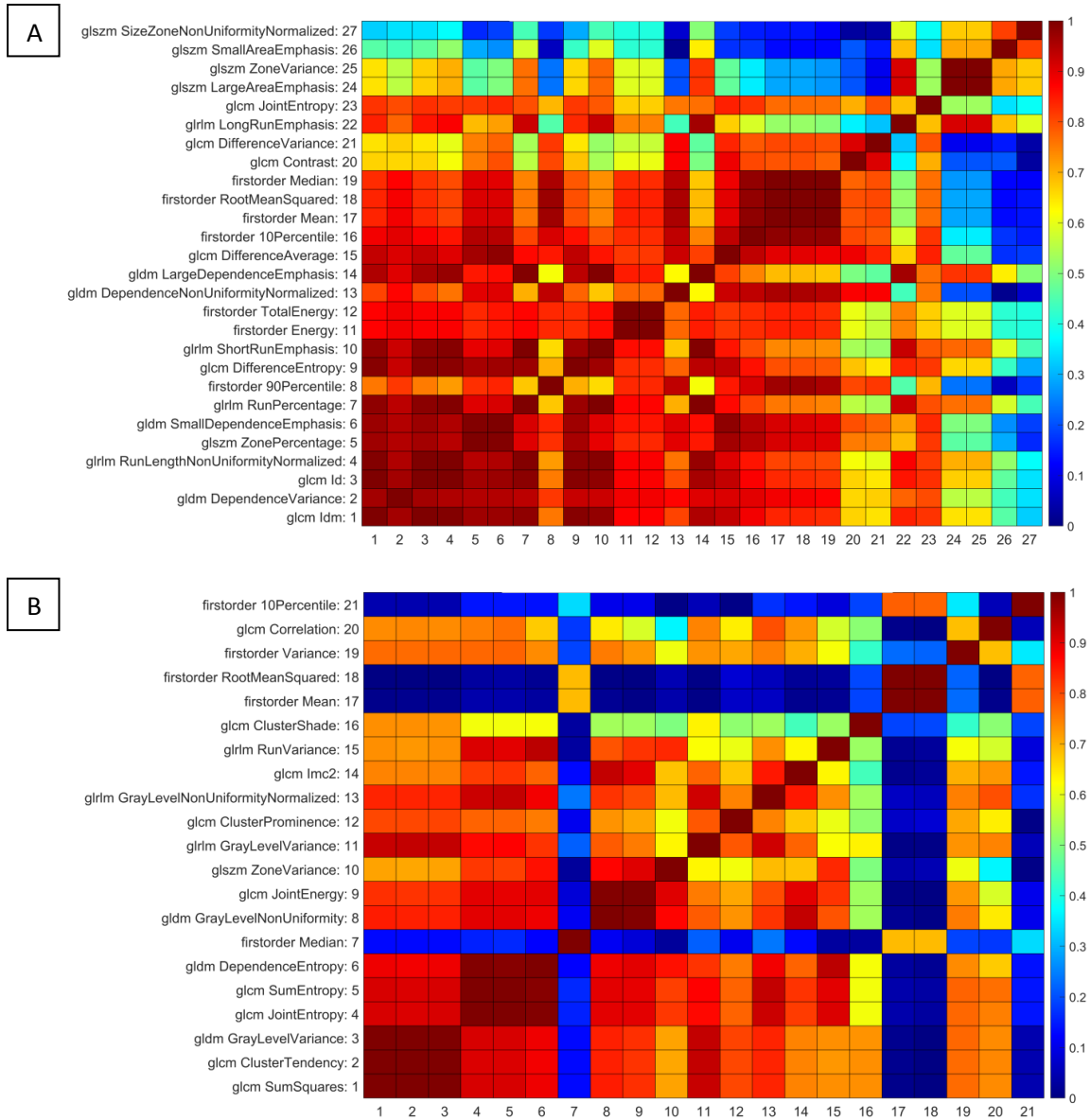
Figure 2.13. The absolute correlation coefficient matrices of SNR-based reduced (A) optical and (B) thermal features for the first patient in the study. The cross-correlation coefficients are indicated in the colour bar, with higher values (red) corresponding to highly correlated features, and lower values (blue) to uncorrelated features.

Features that are highly correlated do not contribute additional useful model training data, cause overfitting of the machine learning model, and hence reduce the accuracy of the classification. To determine and reduce common groups of correlated features among patients, the cross-correlation matrices were discretized (i.e., replacing with binary values of 0, if the absolute correlation value was less than 0.75, and 1 otherwise) and the resulting binary cross-correlation matrices were summed across the patients. The matrices were sorted based on SNR and reordered, placing features of high correlation together for better visualization of dependent groups. The reordered significant correlation coefficients summed over all patients are shown in the matrices in Figure 2.14 (A) and (B) for the SNR-based reduced optical and thermal features, respectively.
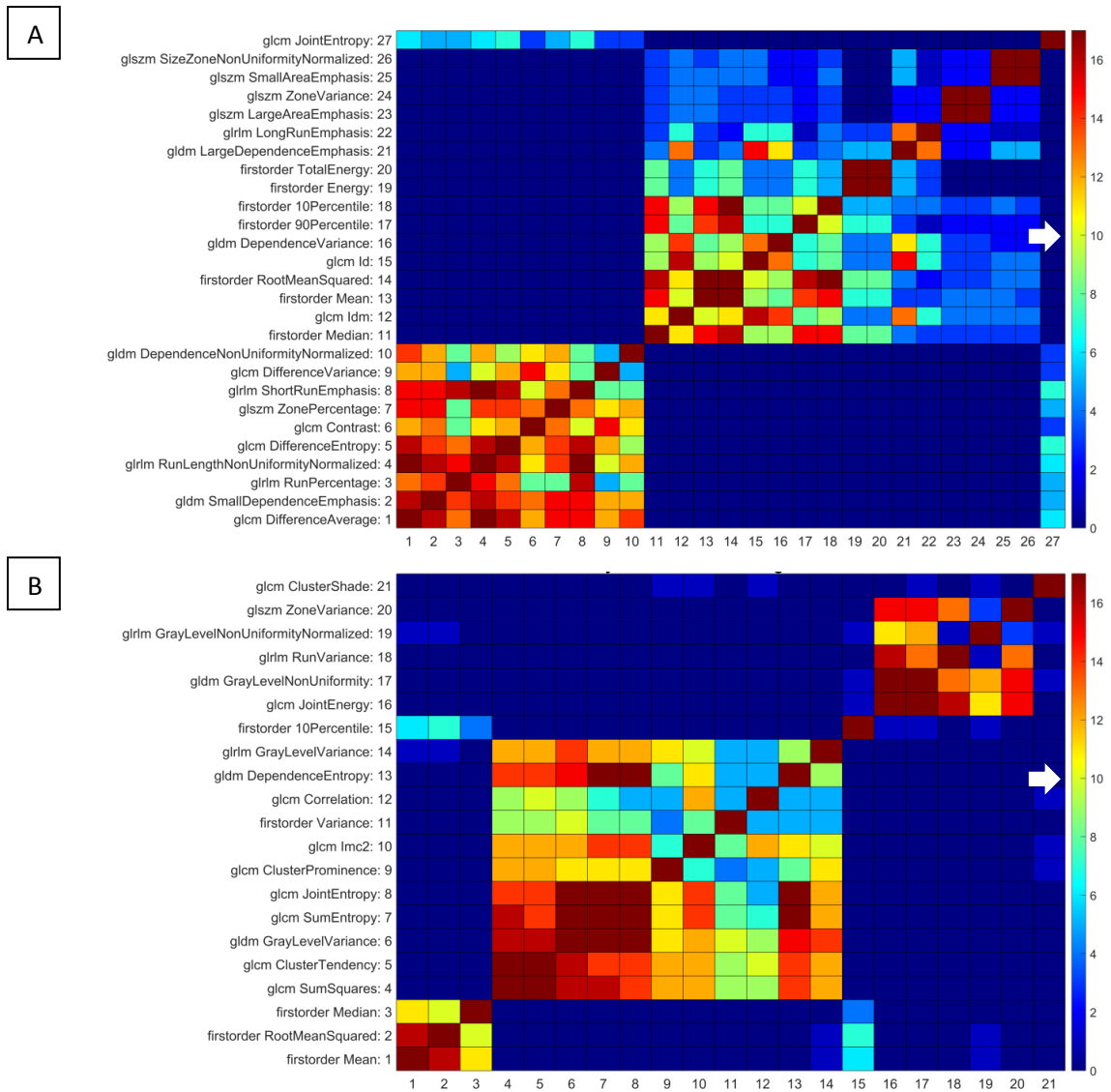
Figure 2.14. The reordered sum of significant correlation coefficient values of SNR-based reduced (A) optical and (B) thermal features. The colour bar indicates the binary correlation value summed over all patients. Features with a coefficient sum criterion of c = 10 or greater were grouped together to identify and eliminate redundancy in the sets.

A significance criterion (c = 10), outlined in Figure 2.14 was used in the summed matrix to group correlated features. Each row (feature) in the matrix was scanned successively to assemble the groups, and the feature with highest SNR was selected from each group. If

features below the row of interest were not added to the group but still correlated ($c \geq 10$) to the highest SNR feature chosen from the group, then they were removed from the matrix to no longer be considered for subsequent grouping. This resulted in a reduced set of independent features. Figure 2.15 depicts the feature grouping and selection process using an example matrix of significant correlation coefficients summed across patients. When scanning the first row, two features exceed the $c = 10$ significance criteria: features 1 and 4. If, hypothetically, feature 4 has the higher SNR, then it will be chosen as the representative feature from that group. Feature 4 is also correlated with feature 3, circled in Figure 2.15, so feature 3 will be removed from the matrix even though it was not considered for the first group. Features 2 and 5 are not correlated with any remaining features so they are inherently selected for groups 2 and 3.
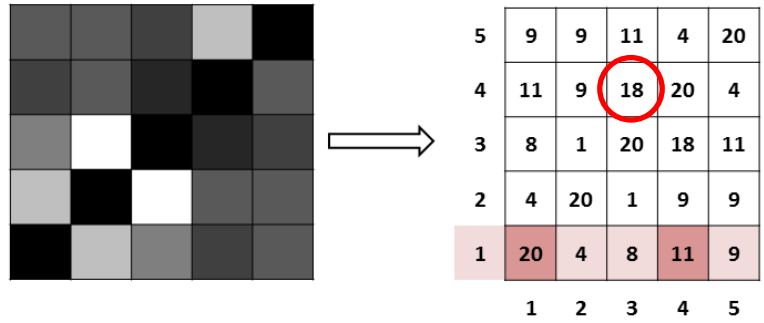
Group 1: Features 1 and 4
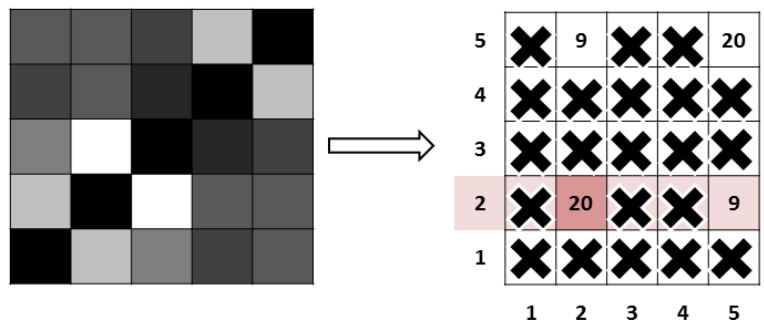
Group 2: Feature 2
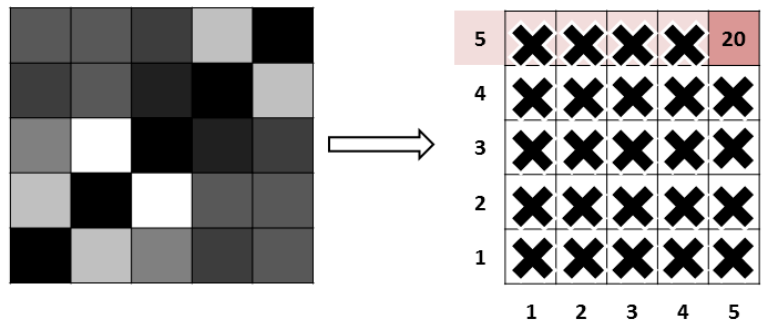
Group 3: Feature 5

Only 3 groups of independent features established.

Figure 2.15 Feature selection flowchart from matrix of reordered significant correlation coefficients summed over all patients.

As a final step in the dimensionality reduction routine, the relationship between reduced feature set and skin toxicity grade were assessed and used to eliminate features with no significant association to skin toxicity. The Kendall Tau correlations between every skin toxicity grade recorded throughout the study, and corresponding change in feature values (with respect to the first therapy session) in the reduced set were computed to validate the selected group of features' correspondence to toxicity, and hence their use in predicting skin toxicity. The Kendall Tau correlation is a non-parametric test, so it does not rely on strict assumptions of the data distribution, or require continuous data. Instead, it assesses the statistical associations between variables based on ranked and numbered data. The dependent variable of our data is ordinal (skin toxicity grade), so this test is valid for the analysis. Kendall Tau correlation is calculated based on concordant and discordant pairs, so it is less sensitive to error and discrepancies in data compared to similar non-parametric rank tests like the Spearman correlation, which is calculated based on deviations.

2.3.4 Toxicity Prediction using Machine Learning Classification

Any features that demonstrated significant correlation to toxicity (based on the Kendall Tau p-values computed for the reduced feature set) were further used in various machine learning algorithms as predictors. Data from the end of the treatment were excluded using BED cut-offs of 50, 45, 40, 35, 30, 25, 20, 15 Gy, as well as no cut-off, to test the predictive capabilities of the images from earlier fractions. This is analogous to a situation where a patient is early on in their treatment and only a few images are available. The linear fit first order coefficients of the selected features (as function of

BED) were used as input data to train and test 25 machine learning models in Matlab's Classification Learner. The Classification Learner is an interactive application that allows a selection of machine learning models to be trained back-to-back, including 3 decision tree models, 2 discriminant analysis models, logistic regression, two naive Bayes models, 6 support vector machines, 6 nearest neighbor kernel approximation models, 5 ensembles, and 3 neural networks. Once the training is complete, the accuracy scores, prediction performance and ROC curve among all models can be compared within the application, and the fitted model can be exported to a Matlab file.

The maximum CTCAE skin toxicity grade that the patients experienced throughout the study was used as response variable in the machine learning training. Classes were constructed by converting the CTCAE grades into binary data: patients who experienced no skin reactions to low grade toxicity (CTCAE ≤ 1) were classified as zero, and those who experienced moderate to high grade toxicity (CTCAE ≥ 2) were classified as one. The data was partitioned into k = 4 groups for cross-validation testing. K-fold is a technique that splits data into k equal groups and uses one of the groups from the data as a test set, and the model is trained on the remaining k-1 groups. Next, the test set observations are fed to the trained model, and as an output, the corresponding predicted class labels are returned along with their probabilities of being classified either as zero or one. The accuracy is computed for that test set by dividing the number of correct predictions by the number of total predictions, defined by equation (11).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} * 100\% \qquad (11)$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

The model then is discarded, and the above procedure is repeated for the remaining subsets. The cross-validation accuracy is defined as the average accuracy across all k-folds.

Although accuracy is a simple metric to compute and understand, it might not be the best metric to evaluate the performance of the model if the data are imbalanced (e.g., more negative classes than positive). The receiver operating characteristic (ROC) curve provides a broader measure of the classifier's performance since it considers all possible discrimination probability thresholds and evaluates the sensitivity and specificity of class prediction separately. Sensitivity is defined as the probability of correctly identifying patients with a disease, while specificity is the probability of correctly identifying patients without a disease. To produce an ROC curve, the True Positive Rate (TPR), or sensitivity, and False Positive Rate (FPR) or (1-speicificity), are computed (equations (12) and (13*)).

$$\text{TPR} = \text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (12)$$

$$\text{FPR} = 1 - \text{Specificity} = \frac{\text{FP}}{\text{FP} + \text{TN}} \qquad (13)$$

An ROC curve is produced by plotting TPR as a function of FPR at different classification thresholds. From the ROC curve, the Area Under the Curve (AUC) provides a measure of the model's performance. The higher the AUC value, the higher the probability that the model will rank a random positive case more highly than a random negative case (i.e., AUC = 1 would indicate perfect discrimination of classes regardless of the chosen threshold). Similar to the computation for accuracy, the AUC is calculated for each fold, and the mean value across all folds is the cross-validated AUC estimate.

The Support Vector Machine (SVM) model sustained the highest cross-validation accuracy throughout all BED cut-offs compare to the other models tested in the Classification Learner. SVM uses a subset of points (support vectors) to maximize the margin of the hyperplane separating classes. It is a justifiable classifier in our case because it performs well for data sets with a small sample and larger number of features. It is also robust to outliers since the margin separating classes depends only on a subset of points. We evaluated the performance of three SVM models (linear, quadratic, and cubic SVMs) using ROC curves and corresponding AUC scores. The cost parameters for the SVM models were tuned through optimization within the SVM algorithm.

Chapter 3   Results

3.1   Radiomic Feature Redundancy Reduction

The significant correlation values summed across patients, used as a metric to eliminate redundant features, resulted in a set of six optical features; GLCM joint entropy, GLSZM small area emphasis, GLSZM large area emphasis, GLDM large dependence emphasis, first order root mean square, GLCM difference average, and five thermal features; GLCM cluster shade, GLSZM zone variance, first order 10th percentile, GLCM correlation, and first order mean. The definitions of these features from Python's documentation of Pyradiomics are assembled in Table 3.1 [53].

Table 3.1 Definitions of the selected features extracted from the Pyradiomic documentation [53].

Selected Texture Features	Definition
First order root mean square	The square-root of the mean of all the squared intensity values. It is another measure of the magnitude of the image values.
GLCM joint entropy	Measure of the randomness/variability in neighborhood intensity values
GLSZM small area emphasis	Measure of the distribution of small size zones, with a greater value indicative of more smaller size zones and more fine textures
GLSZM large area emphasis	Measure of the distribution of large area size zones, with a greater value indicative of more larger size zones and more coarse textures.
GLDM large dependence emphasis	A measure of the distribution of large dependencies, with a greater value indicative of larger dependence and more homogeneous textures
GLCM difference average	Measures the relationship between occurrences of pairs with similar intensity values and occurrences of pairs with differing intensity values
GLCM cluster shade	A measure of the skewness and uniformity of the GLCM. A higher cluster shade implies greater asymmetry about the mean
GLSZM zone variance	Measures the variance in zone size volumes for the zones
First order 10th percentile	The 10th percentile of the intensity histogram
GLCM correlation	A value between 0 (uncorrelated) and 1 (perfectly correlated) showing the linear dependency of gray level values to their respective voxels in the GLCM
First order mean	The average gray level intensity within the ROI

The sum of significant correlation values of these features are shown in Figure 3.1. The high correlation values only reside along the diagonal, implying that the features in the matrix have been successfully reduced to an independent set, generalized to the patients in our study. The correlation coefficients between thermal and optical features were not analyzed. The thermal and optical features might be correlated to an extent but they can still hold information that differs from one another, as they represent different physical aspects of the skin. If this is the case then we would want to keep the features from both types of images and analyse them separately.
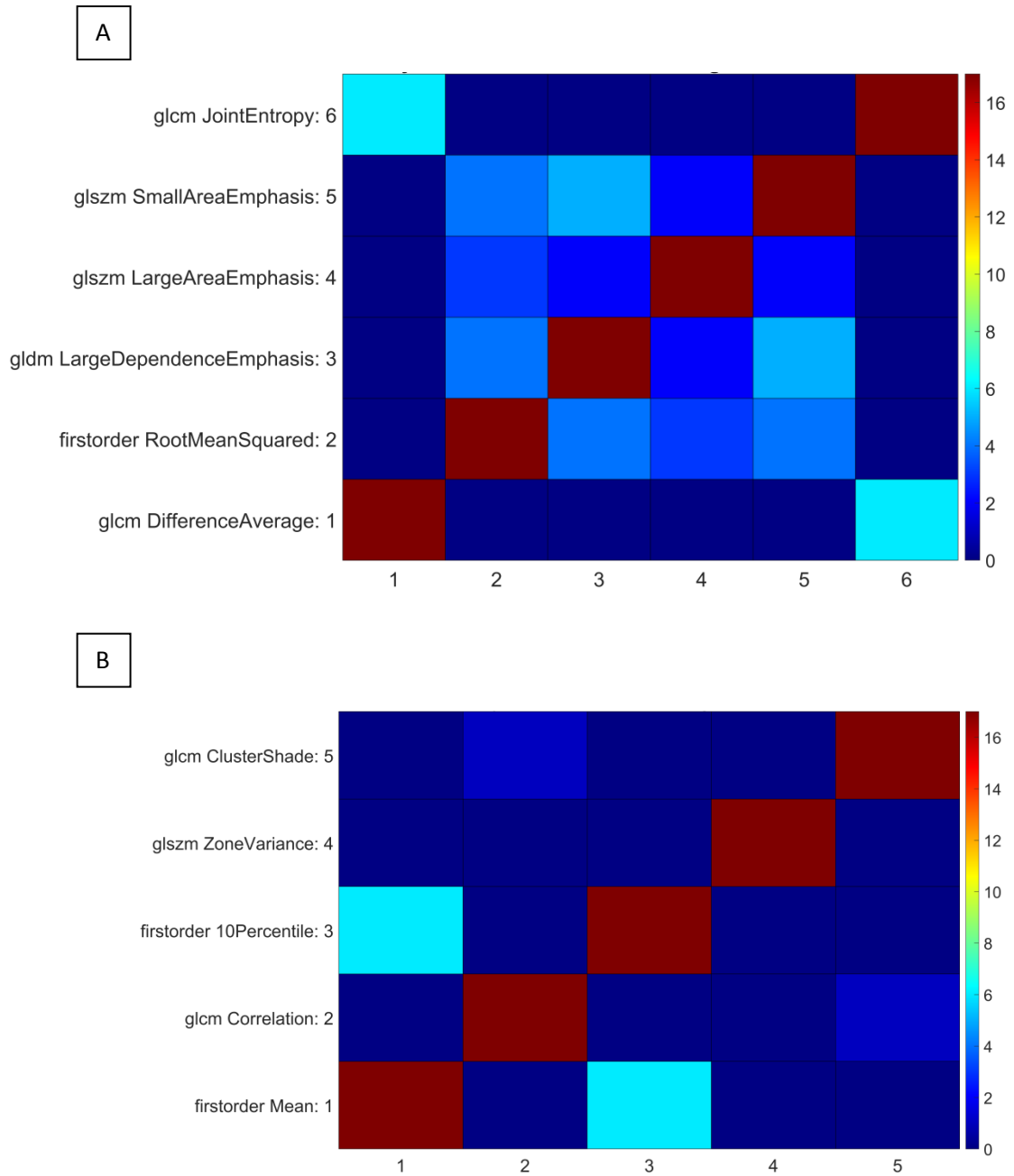
Figure 3.1. The sum of significant correlation coefficient values of (A) optical and (B) thermal features resulting from SNR-based reduction, followed by cross-correlation based reduction. The colour bar indicates the binary correlation value summed across all patients.

The Kendall Tau correlations coefficients and p-values between every skin toxicity grade recorded and the corresponding change in feature values from the initial fraction are presented in Table 3.2 for the reduced six optical and five thermal features. A p-value of $p \leq 0.05$ suggests a statistically significant Kendall Tau correlation.

Table 3.2. The Kendall Tau correlation coefficient and p-values for the 11 SNR- and intrafeature correlation-reduced features. The significant Kendall Tau coefficients and p-values are highlighted in the table.

Radiomic Feature	Kendall Tau Correlation Coefficient	p-value
Optical		
GLCM joint entropy	0.35	7.3×10^{-8}
GLSZM small area emphasis	- 0.27	3.6×10^{-5}
GLSZM large area emphasis	- 0.19	0.003
GLDM large dependence emphasis	- 0.39	1.1×10^{-9}
first order root mean square	- 0.44	6.6×10^{-12}
GLCM difference average	0.39	2.7×10^{-9}
Thermal		
GLCM cluster shade	- 0.031	0.65
GLSZM zone variance	0.16	0.016
first order 10th percentile	0.14	0.032
GLCM correlation	- 0.062	0.33
first order mean	0.24	1.5×10^{-4}

All features except for thermal GLCM cluster shade and thermal GLCM correlation had

significant Kendall Tau p-values (≤ 0.05). Figure 3.2 shows the correlation between the

feature value difference (change in feature value from the first fraction) and

corresponding toxicity grade for all selected features, in the form of box and whisker plots.
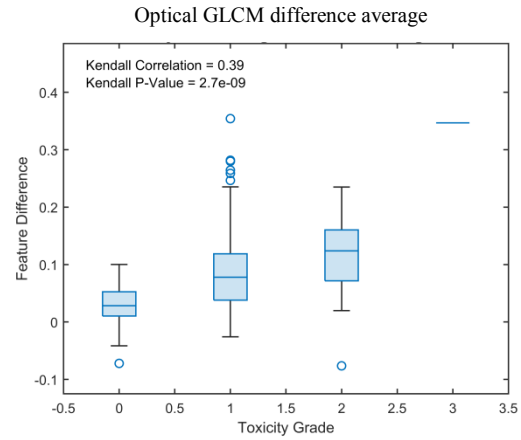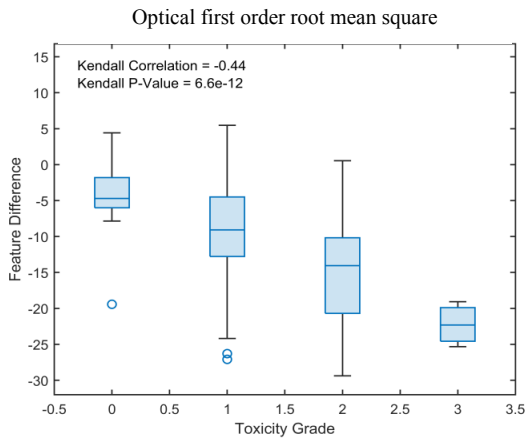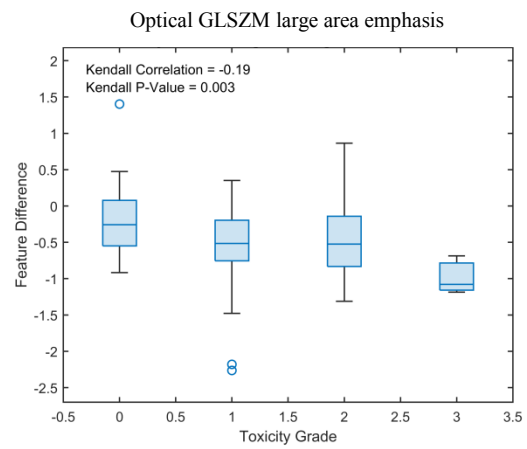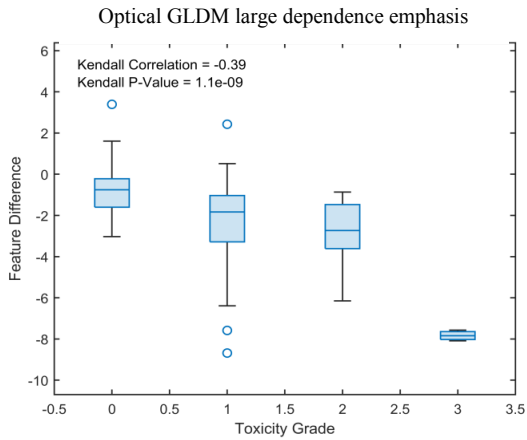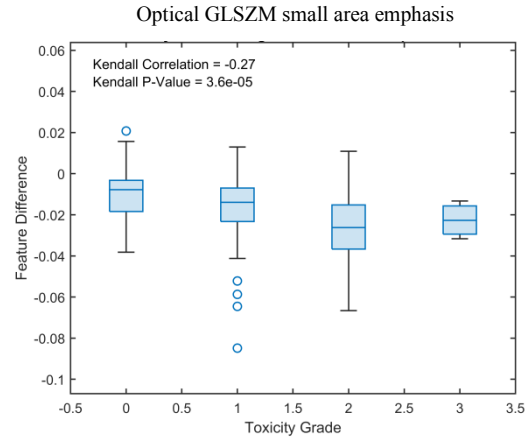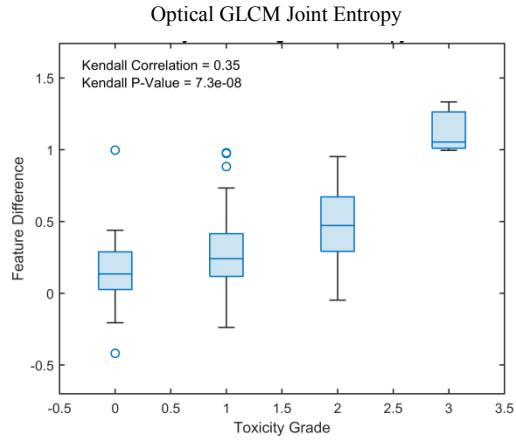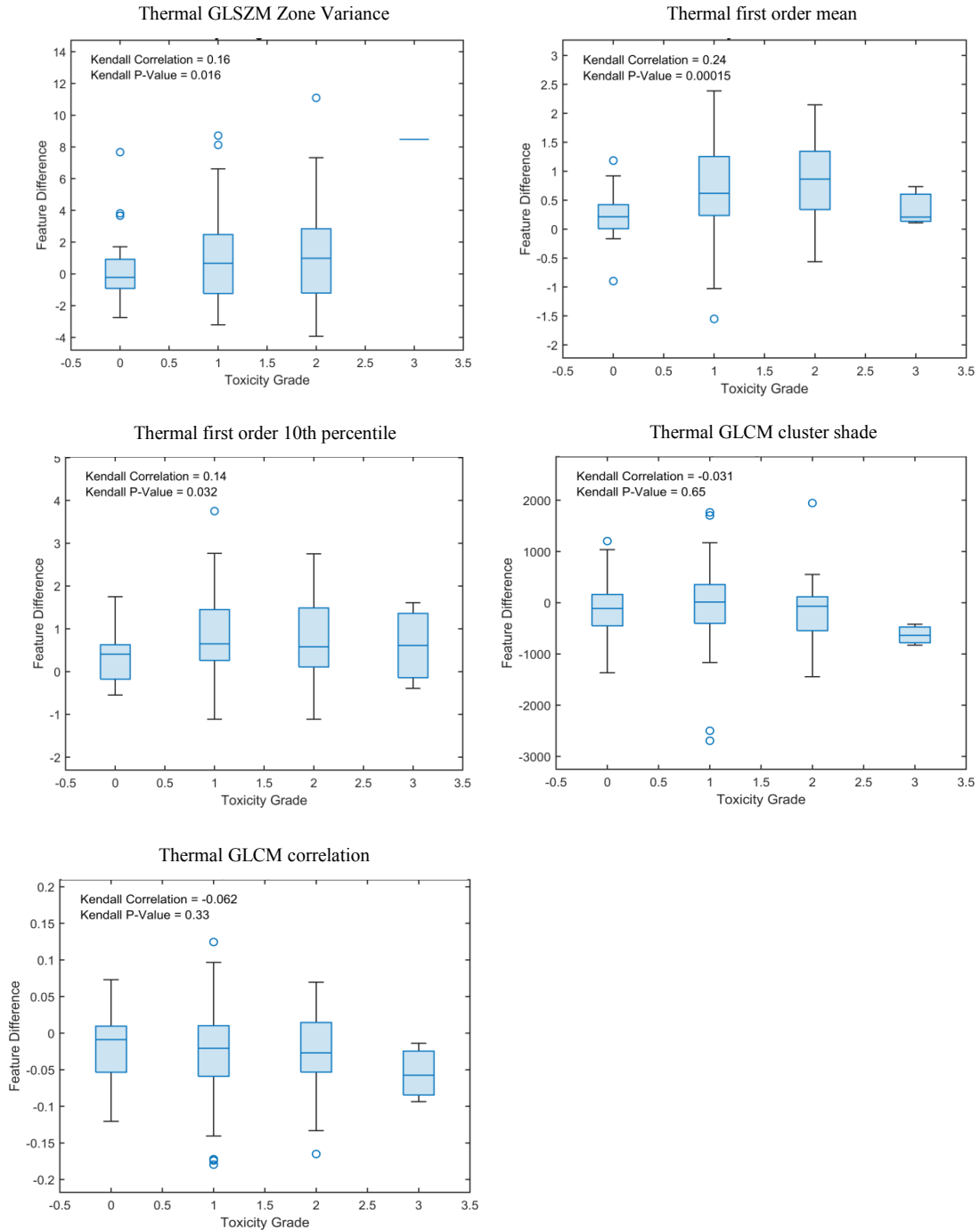
Figure 3.2. Box and whisker plots of feature value differences (taken between the first fraction and each fraction with recorded toxicity) and the corresponding toxicities for all features from the reduced set. The line inside of each box is the feature difference median within that grade. The top and bottom edges of each box correspond to the 0.75 and 0.25 quantile, respectively, and the whiskers correspond to the maximum and minimum values. The outliers (shown in circles) were computed using 1.5 x the interquartile range (0.75 - 0 .25), and were not included in the correlation computation.

The box and whisker plots for the other seven features (not shown) presented a notable correlation between the feature value and corresponding toxicity grade. The thermal GLCM cluster shade and thermal GLCM correlation were removed from the feature set since they did not demonstrate significant correlation, and the remaining six optical and three thermal features (nine in total) were further used as predictors in training the SVM classifiers.

3.2   Skin Toxicity Diagnostics

The three SVM classifiers were trained in five repeated runs (to average over the experiments and obtain errors) for each of the nine BED cut-offs (50, 45, 40, 35, 30, 25, 20, 15 Gy, and no cut-off). This resulted in 135 cross-validation accuracy values and 135 ROCs curves used to evaluate the model's performance. Two sample ROC curves for the linear SVM are displayed in Figure 3.3.
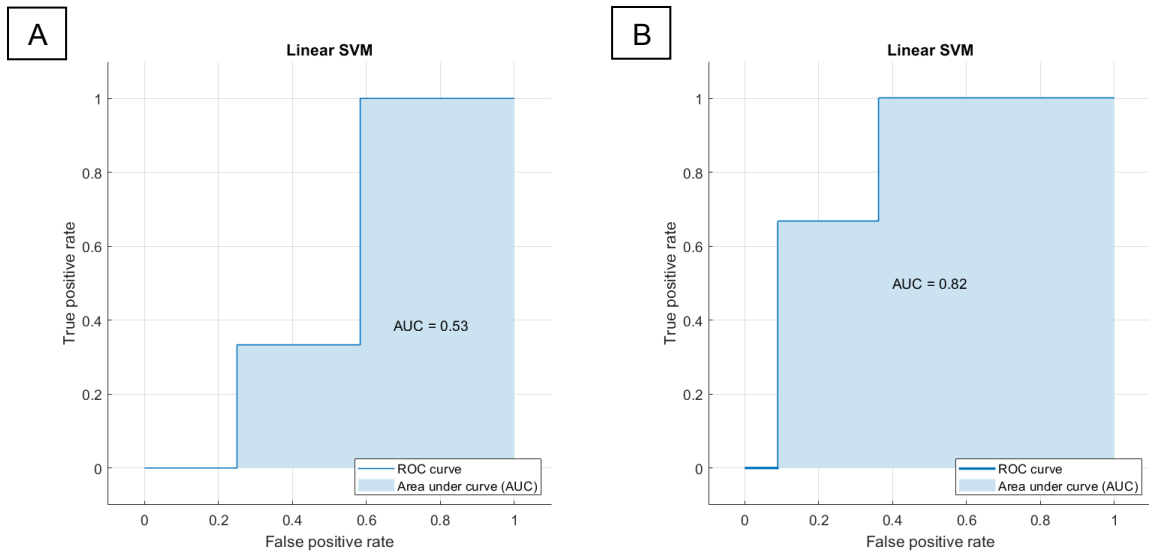
Figure 3.3. ROC curves for one of the five repetition runs of the linear SVM with BED cut-offs of A) 15 Gy and B) 50 Gy. A random classifier would give points on the ROC curve lying along the diagonal, whereas a classifier with good performance would produce a curve closer to the top left corner.

The curves above were chosen as an example because they show the visual difference between poor discrimination and good discrimination in an ROC curve, and the effect on the resulting AUC values. An AUC of 0.53 suggests poor discrimination (Figure 3.3 A), while an AUC of 0.82 (Figure 3.3 B) provides good discrimination between classes.

The accuracy and AUC values were averaged from the five repeated runs, and plotted for the three SVM models as a function of BED cut-off in Figure 3.4. 'No cut-off' was set to 75 Gy as this was the maximum BED delivered in the study.
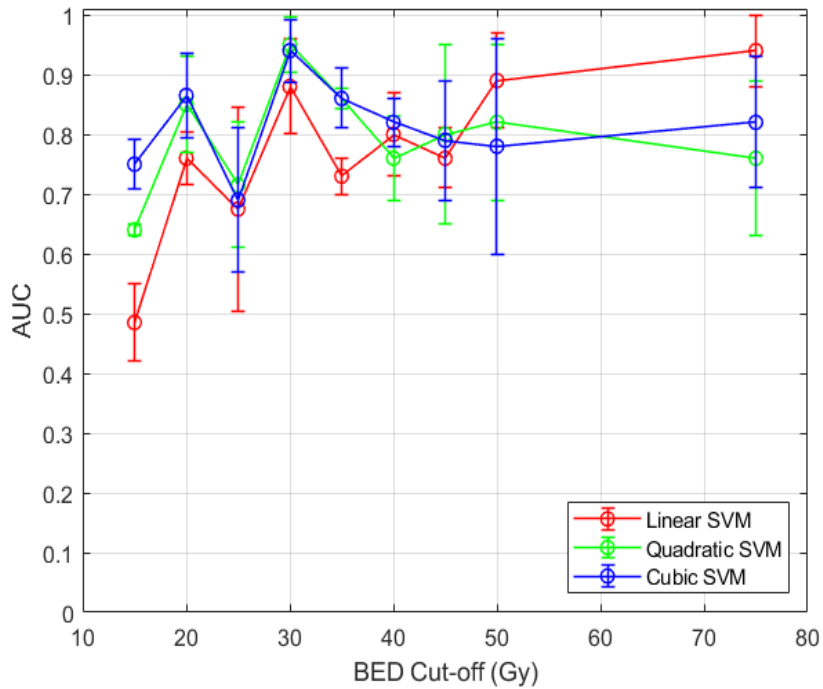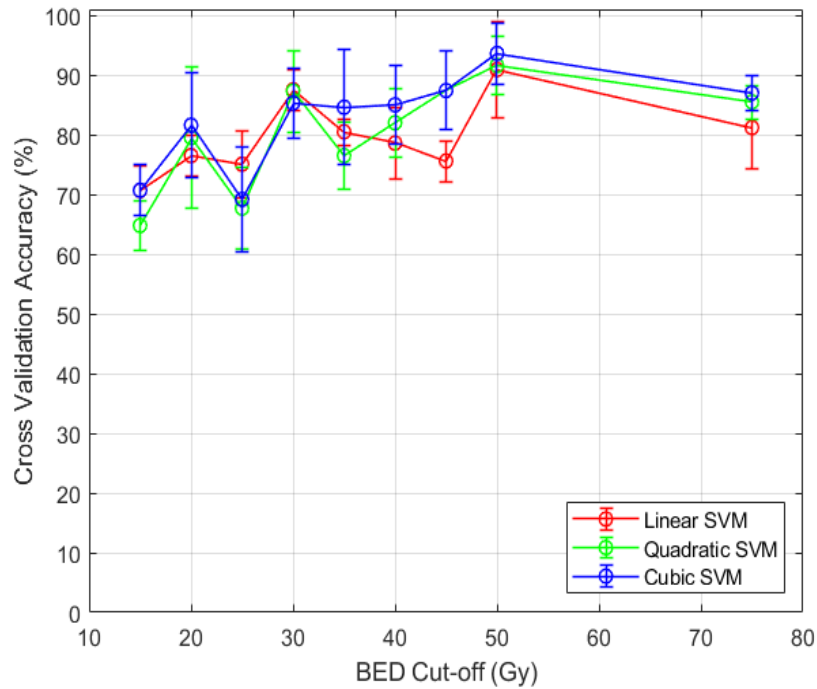
Figure 3.4. Cross-validation (k-fold = 4) linear, quadratic and cubic SVM accuracy and AUC for different image data BED cut-offs, both metrics averaged across five runs. The error bars on the accuracy and AUC are the standard deviation values from the five runs.

Chapter 4   Discussion

4.1   Dependence of model performance on data distribution and signal robustness

A BED-series of 216 radiomic features, extracted from optical and thermal images of the patients' skin were filtered based on the SNR of the feature value-BED linear fit and intra-feature correlation, reducing them to a set of six optical and five thermal statistically dependent features. The Kendall Tau correlation values between the 11 features and toxicity grading were computed to justify the selection of those features for subsequent predictive model fitting. Kendall Tau correlation p-values of the reduced features were significant among all six optical features, but only three out of the five thermal features presented significant correlation to toxicity. A thermal camera with better characteristics, such as sensitivity, accuracy and spatial higher resolution, may allow for more sensitive changes to be detected with respect to the control ROI. So, although two thermal features were not deemed usable for the predictive model, the same routine repeated on images taken with a higher resolution camera would potentially result in a significant correlation to skin toxicity grade, and therefore avoid rejection of any of the SNR-based and intra-feature correlation reduced features.

Twenty-five different classification models were tested to choose those providing most accurate prediction. Out of all tested models, the linear, quadratic and cubic SVMs demonstrated the best performance across a range of BED cut-offs and their cross-validation and AUC values were further analyzed and reported in this study. Although the accuracy is a simple metric to interpret, it is misleading for unbalanced data where the sample distribution is skewed more towards either the positive or negative class. Like in

our case, throughout the entire course of the study, only 6 patients experienced minimal symptoms (CTCAE $\leq$ 1) while 14 patients faced moderate to severe symptoms (CTCAE $\geq$ 2). If, for example, a model is biased towards patients with CTCAE $\geq$ 2 and all patients with CTCAE $\leq$ 1 are misclassified, then it will not affect the accuracy in the same way as an evenly distributed sample. The AUC metric sweeps across different classification thresholds and evaluates sensitivity as a function of specificity, i.e., two metrics computed from the positive cases and negative cases separately. Therefore, even the class with the minority of cases will have a strong impact on the AUC value. This method is more appropriate for the analysis for our data because it is invariant to class distribution.

To test predictability, robustness and representability of the SVM model, images were excluded from the end of the treatment using various BED cut-offs. The quadratic and cubic SVM showed superior predictive performance with cross-validation accuracy remaining above 80%, and the AUC stayed above 0.7 for all three models when reducing the input to 30 Gy BED, or equivalently, the first third or first half of treatment fractions, depending on the patient's fractionation schedule. The AUC for the linear SVM outperformed the quadratic and cubic SVMs for no-cutoff and a 50 Gy cutoff, but as the more data was discarded from the end of treatment, the AUC of the quadratic and cubic SVMs increased. This could be due to the fact that the decision margin of the quadratic and cubic SVM is a polynomial function rather than straight line. Adding adjustable curvilinearity to the decision line allows producing a higher level of complexity of margins separating classes of the support vectors, when fitted with polynomial functions, therefore allowing a smaller margin. When computing the ROC metrics, the classification threshold is varied, so the probability scores of a tightly fit separation margin are more

likely to be misclassified. However, when points from the end of the feature-BED data are removed, there is more variability and uncertainty in the feature-BED trend line first order coefficients used as the input for the classification model. This would result in a more generalized and larger margin, with a more forgiving decision line. Therefore, we would expect fewer misclassifications when varying the threshold.

Saedina et. al reported a test accuracy of 0.87, and test AUC of 0.98 for a random forest classifier using data from the fifth fraction (one third way through the course of treatment), however the error on these values were not reported [35]. Figure 3.4 shows margins of error as large as 15% for the test accuracy and 30% for the test AUC in our data, as well as a considerable margin of error variation at different fractions. This result signifies the importance of training the model at a number of different points throughout the treatment, and evaluating the model's error when repeating runs for different partitions. Random variations in the data would have a larger effect on the prediction outcome if we only used images from one fraction to make predictions rather than the first order linear coefficients of the features over several fractions. The larger sample size in the Saedina et. al study (n=90) [35] could explain the superior accuracy and AUC compared with ours (from n=20). One of the strengths and unique aspects of this study is that the image acquisition used simple tools that could be used by the implemented at the point of care, with the patient in treatment position. This is in contrast to previous work [35] that required a separate room and setup for patients. The methodology integrated smoothly into the workflow, allowing for acquisition of optical and thermal image data on every fraction.

We believe that improvements in our results would be possible with regard to the thermal imaging, with thermal features and metrics limited by the spatial resolution of the images. The FLIR One Pro did not have a high enough resolution in its thermal images to detect meaningful changes in higher order ipsilateral-contralateral features. To properly represent the details of skin structure when evaluating higher order features, a sufficient number of pixels covering the ROI and affected areas is required. The fewer the number of pixels, the coarser the level of detail in representation of the skin structure, adding uncertainty to the higher order features, so those metrics were only assessed from the ipsilateral side alone. This does not preclude using thermal images in future, but either a better camera with higher sensitivity, accuracy and spatial resolution, needs to be used, or the radiomic features adjusted to account for low resolution.

Chapter 5   Conclusion

5.1   Summary of Work

One of the most recognized challenges in the field in radiation therapy is to limit damage to normal tissues in order to minimize toxicity, without compromising the success of the treatment in terms of tumor control. However, response to radiation therapy shows large variability among patients, with dependence on many factors, including treatment parameters and patient specific details, making it difficult to determine the best curative approach.

The 20 patients who were included in our study experienced a wide range of radiation skin reactions. For each patient, the differences in thermal and optical image properties between irradiated and unirradiated sides were analyzed as a function of increasing dose, with each patient serving as their own control. The number of optical and thermal radiomic features from these patients' images obtained as part of the initial analysis was automatically reduced using a procedure based on the SNR and cross-correlation feature sorting methods. The relationship between these sets of features and skin toxicity were analyzed using Kendall Tau correlation, and corresponding p-values were used to reduce the series further. The final set of features were trained in machine learning models to assess the predictive capabilities of optical and thermal images at early stages of treatment.

A total of 676 images allowed effective training of a SVM; linear, quadratic and cubic SVMs were capable of providing an accuracy above 80%, and an AUC above 0.7 after imaging just 30 Gy BED, ranging from as few as one third to one half of the total treatment fractions. The results provided firm support for the association of skin toxicity with the selected radiomic features at the earlier stages of treatment.

5.2   Future Work

There were few lessons learned from the analysis of limitations and ways of further improvement experienced through the course of this project. Limitations can be subdivided in two categories; concerns the overall patient sample size and representation, and technical challenges encountered during the project.

With respect to the first category, one of the major limitations of this study is that it did not include a wide range of different skin colours. We also did not evaluate diabetes, smoking habits, and age as separate variables. The overall sample size was expected to be larger when planning the project, however, the restriction imposed by the COVID-19 regulations made recruitment more problematic, challenging and overall slower than before the pandemic. Under these circumstances, even reaching 20% of the expected sample size was considered as a success by our research team. We strongly encourage a continuation of the study of various skin reactions on radiation treatment and prediction of skin toxicity using objective methods. The next step in this direction would be to include a larger and more diverse group of patients in any study similar to ours to capture the effects of various patient-related factors and skin characteristics on treatment outcome with a larger sample size. Such a continued study could be made multi-institutional by using identical equipment and unified procedures in two or more clinics as part of a joint collaborative study. This would allow development and testing of a more robust predictive algorithm than if basing all the work on data from a single clinic.

With respect to the technical limitations that can likely be addressed in the future, while being satisfied with the optical image quality provided by a front-of-the-line camera, we found that the performance of the thermal imaging could be improved to allow a more comprehensive radiomic analysis.

Overall, quantitative analysis of radiomic features acquired from optical and thermal images was shown to yield significant biomarkers for detecting and predicting skin toxicities. Future work will continue toward the development of a practical and efficient tool that can be used daily in the treatment room to predict skin toxicity.

Bibliography

[1]   A. Patel, "Benign vs Malignant Tumors," *JAMA Oncol,* vol. 6, no. 9, p. 1488.

[2]   Goverment of Canada, "Canadian Cancer Statistics 2021," Health Promotion and Chronic Disease Prevention in Canada, 7 February 2022. [Online]. Available: https://doi.org/10.24095/hpcdp.41.11.09.

[3]   Canadian Cancer Society, "Breast Cancer Statistics," 2022. [Online]. Available: https://cancer.ca/en/cancer-information/cancer-types/breast/statistics.

[4]   R. Baskar, K. A. Lee , R. Yeo and K. W. Yeoh, "Cancer and radiation therapy: current advances and future directions," *Int J Med Sci,* vol. 9, no. 3, pp. 193-199, 2012.

[5]   R. M. Jordan and J. Oxenberg, "Breast Cancer Conservation Therapy," 2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK547708/.

[6]   M. Teoh, C. H. Clark, K. Wood, S. Whitaker and A. Nisbet, "Volumetric modulated arc therapy: a review of current literature and clinical use in practice," *Br J Radiol,* vol. 84, no. 1007, pp. 967-996.

[7]   T. Virén, J. Heikkilä and K. Myllyoja, "Tangential volumetric modulated arc therapy technique for left-sided breast cancer radiotherapy," *Radiat Oncol,* vol. 10, no. 79, 2015.

[8]   D. Karpf, M. Sakka, M. Metzger and G. Gerhard, "Left breast irradiation with tangential intensity modulated radiotherapy (t-IMRT) versus tangential volumetric modulated arc therapy (t-VMAT): Trade-offs between secondary cancer induction risk and optimal target coverage," *Radiation Oncology,* vol. 14, no. 1, 2019.

[9]   Canadian Cancer Society, "The Lymphatic System," 2022. [Online]. Available: https://cancer.ca/en/cancer-information/what-is-cancer/lymphatic-system.

[10]   M. Rahman and S. Mohammed, "Breast cancer metastasis and the lymphatic system," *Oncol Lett,* vol. 10, no. 3, pp. 1233-1239, 2015.

[11]   J. D. Brierley, M. K. Gospodarowicz and C. Wittekind, TNM Classification of Malignant Tumours. 8th ed., Wiley Blackwell, 2017.

[12]   R. Devi, G. Anandhamala and R. Sundhar, "Recent Trends in Medical Imaging
        Modalities and Challenges For Diagnosing Breast Cancer," *Biomedical and
        Pharmacology Journal,* vol. 11, no. 3, pp. 1649-1658, 2018.

[13]   S. B. Edge, D. R. Byrd, C. C. Compton and A. G. Fritz, AJCC cancer staging
        manual, 7th edition, Springer, 2010.

[14]   J. Koh and M. J. Kim, "Introduction of a New Staging System of Breast Cancer
        for Radiologists: An Emphasis on the Prognostic Stage," *Korean J Radiol,* vol.
        20, no. 1, pp. 69-82, 2019.

[15]   Canadian Cancer Society, "Stages of Breast cancer," 2022. [Online]. Available:
        https://cancer.ca/en/cancer-information/cancer-types/breast/staging.

[16]   L. L. Gunderson and J. E. Tepper, Clinical Radiation Oncology (2nd ed.),
        Philadelphia: Elsevier, 2006.

[17]   BC Cancer, "Radiation Therapy," 2022. [Online]. Available:
        http://www.bccancer.bc.ca/books/breast/management/breast-cancer-in-pregnancy.

[18]   A. M. Brunt, J. S. Haviland, A. M. Kirby, N. Somaiah, J. R. Yarnold, D. A.
        Wheatley and J. M. Bliss, "Five-fraction Radiotherapy for Breast Cancer: FAST-
        Forward to Implementation," *Clinical Oncology,* vol. 33, no. 7, 2021.

[19]   C. F. Njeh, M. W. Saunders and C. M. Langton , "Accelerated Partial Breast
        Irradiation (APBI): A review of available techniques," *Radiation Oncology,* vol.
        5, no. 90, 2010.

[20]   H. Mursh, Fundamentals of Radiation Oncology: Physical, Biological, and
        Clinical Aspects, 2019.

[21]   J. R. Herndon, "Are Radiation Boosts Necessary?," verywellhealth, [Online].
        Available: https://www.verywellhealth.com/what-is-a-radiation-boost-for-breast-
        cancer-
        430383#:~:text=The%20radiation%20boost%20may%20be,the%20primary%20t
        umor%20was%20located..

[22]   S. Yao, Y. Zhang and K. Nie, "Setup uncertainties and the optimal imaging
        schedule in the prone position whole breast radiotherapy," *Radiat Oncol,* vol. 14,
        no. 76, 2019.

[23] J. W. Leer, "What the clinician wants to know: radiation oncology perspective. Cancer Imaging.," vol. 5, 2005.

[24] A. S. Saini, C. S. Hwang, M. C. Biagioli and I. J. Das, "Evaluation of sparing organs at risk (OARs) in left-breast irradiation in the supine and prone positions and with deep inspiration breath-hold," *J Appl Clin Med Phys,* vol. 19, no. 4, pp. 195-204, 2018.

[25] National Cancer Institute (U.S.), "Common terminology criteria for adverse events: (CTCAE)," 2017.

[26] J. L. Wright, C. Takita, W. Zhao and I. M. Reis, "Prospective evaluation of radiation-induced skin toxicity in a race/ethnically diverse breast cancer population," *Cancer medicine,* vol. 5, no. 3, p. 454–464, 2016.

[27] J. P. Pignol, I. Olivotto, E. Rakovitch, S. Gardner S, K. Sixel , W. Beckham, T. T. Vu, P. Truong , I. Ackerman and L. Paszat, "A Multicenter Randomized Trial of Breast Intensity-Modulated Radiation Therapy to Reduce Acute Radiation Dermatitis," *J Clin Oncol,* vol. 26, no. 13, p. 2085–2092, 2008.

[28] R. Suresh, J. Raffi, F. Yuen and J. Murase, "Treatment of moist desquamation for patients undergoing radiotherapy.," *Int J Womens Dermatol,* vol. 5, no. 2, pp. 124-125, 2019.

[29] C. Lilla, C. Ambrosone, S. Kropp and I. Helmbold, "Predictive factors for late normal tissue complications following radiotherapy for breast cancer," *Br Canc Res Treat,* vol. 106, pp. 143-50, 2007.

[30] J. P. Pignol, G. Mitera, s. Bosnic, T. T. Vu and H. Verkooijen, "Prospective Evaluation of Severe Skin Toxicity and Pain During Postmastectomy Radiation Therapy," *Int J Radiation Oncol Biol Phys,* vol. 91, pp. 157-164, 2015.

[31] M. F. Chen, W. C. Chen, C. H. Lai and C. H. Hung, "Predictive factors of radiation-induced skin toxicity in breast cancer patients," *BMC cancer,* vol. 10, no. 508, 2010.

[32] A. Bostrom, "Potent corticosteroid cream (mometasone furoate) significantly reduces acute radiation dermatitis: results from a double-blind, randomized study," *Radioth Oncol,* vol. 59, pp. 257-265, 2001.

[33]   O. Maillot, N. Leduc, V. Atallah and P. Escarmant, "Evaluation of acute skin toxicity of breast radiotherapy using thermography: Results of a prospective single-centre trial," *Cancer Radiother,* vol. 22, no. 3, pp. 205-210, 2018.

[34]   A. Templeton, J. Chu, M. Sun, R. Yao and J. Sun, "Thermal Effusivity Changes as a Precursor to Moist Desquamation," *Radiation Research,* vol. 178, no. 4, p. 295–303., 2012.

[35]   K. Saednia , S. Tabbarah, A. Lagree, T. Wu, J. Klein , E. Garcia, M. Hall , E. Chow, E. Rakovitch, C. Childs, A. Sadeghi-Naini and W. T. Tran, "Quantitative Thermal Imaging Biomarkers to Detect Acute Skin Toxicity From Breast Radiation Therapy Using Supervised Machine Learning," *Int J Radiat Oncol Biol Phys,* vol. 106, no. 5, pp. 1071-1083, 2020.

[36]   N. Mansurov, "What is a DSLR (Digital SLR) Camera?," 2018. [Online]. Available: https://photographylife.com/what-is-a-dslr.

[37]   Astropix, "How Digital Cameras Work," [Online]. Available: https://www.astropix.com/html/astrophotography/how.html.

[38]   Hamamastu, "Concepts in Digital Imaging Technology," [Online]. Available: https://hamamatsu.magnet.fsu.edu/articles/digitalimagebasics.html.

[39]   Teledyne FLIR, "How do thermal cameras work," 18 June 2020. [Online]. Available: https://www.flir.ca/discover/rd-science/how-do-thermal-cameras-work/.

[40]   "https://web.stanford.edu/class/cs101/image-1-introduction.html," [Online].

[41]   V. Parekh and M. A. Jacobs, "Radiomics: a new application from established techniques," *Expert Rev Precis Med Drug Dev,* vol. 1, no. 2, pp. 207-226, 2016.

[42]   S. S. Yip and H. J. Aerts, "Applications and limitations of radiomics," *Phys Med Biol,* vol. 61, no. 13, pp. 150-166, 2016.

[43]   M. E. Mayerhoefer, A. Materka, G. Langs, I. Häggström, P. Szczypiński, P. Gibbs and G. Cook, "Introduction to Radiomics," *J Nucl Med,* vol. 61, no. 4, pp. 488-495, 2020.

[44]   R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang and J. P. Campbell, "Introduction to Machine Learning, Neural Networks, and Deep Learning,"

Trans. Vis. Sci. Tech, vol. 9, no. 2, 2020.

[45]   S. Badillo , B. Banfai, F. Birzele and a. et., "An Introduction to Machine Learning," *Clin Pharmacol Ther,* vol. 107, no. 4, pp. 871-885, 2020.

[46]   Tarang Shah, "About Train, Validation and Test Sets in Machine Learning," 2017. [Online]. Available: https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7.

[47]   Dhiraj K, "Top 4 advantages and disadvantages of Support Vector Machine or SVM," 2019. [Online]. Available: https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107.

[48]   R. Pupale, "Support Vector Machines(SVM) — An Overview," [Online]. Available: https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989.

[49]   "Lab: Support Vector Machines," [Online]. Available: https://hastie.su.domains/ISLR2/Labs/Rmarkdown_Notebooks/Ch9-svm-lab.html.

[50]   Shivam Sharma, "SVM: What makes it superior to the Maximal-Margin and Support Vector Classifiers?," 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/05/support-vector-machines/.

[51]   B. Marples, "The Management of Gynecologic Cancers: Radiobiology," 2013. [Online]. Available: https://www.astro.org/uploadedFiles/Main_Site/Meetings_and_Events/2013_Spring_Refresher_Course/Meeting_Program/RADIOBIOLOGY%20GYN%20MARPLES.pdf.

[52]   B. Emami, G. Woloschak and W. Small, "Beyond the linear quadratic model: intraoperative radiotherapy and normal tissue tolerance," *Translational Cancer Research,* vol. 4, no. 2, 2015.

[53]   Pyradiomics, "Radiomic Features," [Online]. Available: https://pyradiomics.readthedocs.io/en/latest/features.html.

[54]   S. V. Sree, E. Y. Ng, R. U. Acharya and O. Faust, "Breast imaging: A survey," *World J Clin Oncol,* vol. 2, no. 4, pp. 171-178, 2011.

[55]   C. F. Njeh, M. W. Saunders and C. M. Langton, "Accelerated Partial Breast

Irradiation (APBI): A review of available techniques.," *Radiat Oncol,* vol. 5, no. 90, 2010.

[56]   Mathworks, "MattesMutualInformation," [Online]. Available: https://www.mathworks.com/help/images/ref/registration.metric.mattesmutualinf ormation.html.

[57]   R. Berwick, [Online]. Available: https://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf.