Characterization of Gynecological Tumors using Texture Analysis in the Context of an

^{18}F-FDG Adaptive PET Protocol

by

Jeff Wade Nawrocki

Graduate Program in Medical Physics
Duke University

Date:_____

Approved:

_____
Oana Craciunescu, Supervisor

_____
Junzo Chino

_____
Timothy Turkington

Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science in the Graduate Program
in Medical Physics in the Graduate School
of Duke University

2015

ABSTRACT

Characterization of Gynecological Tumors using Texture Analysis in the Context of an ^{18}F-FDG Adaptive PET Protocol

by

Jeff Wade Nawrocki

Graduate Program in Medical Physics
Duke University

Date:_____

Approved:

_____
Oana Craciunescu, Supervisor

_____
Junzo Chino

_____
Timothy Turkington

An abstract of a thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science in the Graduate Program
in Medical Physics in the Graduate School
of Duke University

2015

Abstract

In radiation oncology, ^{18}F-FDG Positron Emission Tomography (PET) is used for determining metabolic activity of cancers as well as delineating gross tumor volumes (GTV) for treatment planning. More recently, PET is being utilized for adaptive therapies for gynecological malignancies in which tumor response may be estimated and treatments adjusted during the course of radiation. In addition to treatment assessment, ^{18}F-FDG PET has become a tool in the prediction of tumor response because of the derived Standard Uptake Value (SUV), a measure of the metabolic activity of a tumor. In this study, we seek to establish texture analysis as complimentary to SUV for predicting tumor response as well as understanding temporal changes during treatment in gynecological cancers. An additional experiment was performed studying the variability of texture features from baseline and intra-treatment PET scans due to reconstruction parameters in order to identify features that show statistically significant changes during treatment and that are independent of reconstruction parameters.

In this IRB approved clinical research study, 29 women with node positive gynecological malignancies visible on PET including cervical, endometrial, vulvar, and vaginal cancers are treated with radiation therapy. Prescribed dose varied between 45-50.4Gy, with a 55-70Gy boost to the PET positive nodes. A baseline, intra-treatment (between 30-36Gy), and post-treatment PET-CT were obtained with tumor response

determined by a physician according to post-treatment RECIST. All volumes were re-contoured on the intra-treatment PET-CT. Primary GTVs were segmented both with the 40% SUV_{max} threshold method and a validated gradient-based contouring tool, PET Edge (MIM Software Inc., Cleveland, OH). A MATLAB Graphical User Interface (GUI) called Duke FIRE (Functional Imaging Research Environment) was developed for this study in order to calculate four mathematical algorithms representing the spatial distribution of pixels in an image: gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), and the neighborhood gray level difference matrix (NGLDM). Features representing characteristics of the image are derived from these texture matrices: 12 local features from the GLCM, 11 regional features from the GLRLM, 11 regional features from the GLSZM, and 5 local features from the NGLDM. Additionally, 6 global SUV histogram features including SUV_{mean}, SUV_{median}, SUV_{max}, skewness, kurtosis, and variance as well as metabolic volume (MV) and total lesion glycolysis (TLG) are extracted. The prognostic power of each baseline feature derived from both gradient-based and threshold segmentation methods was determined using the Wilcoxon rank-sum test. Receiver operating characteristic (ROC) curves were calculated to understand the sensitivity and specificity of baseline texture features compared to SUV metrics. Changes in features from baseline to intra-treatment PET-CT were determined using the Wilcoxon signed-rank test. A subset of 7 patient baseline and intra-treatment raw PET

data was reconstructed 6 times using a TrueX+TOF algorithm on a Siemens Biograph mCT with varying iterations and Gaussian filter widths. Texture features were derived from the GTV as before. Texture features per patient were normalized to the respective clinical baseline value in order to limit variability to reconstruction parameters. Mean percent ranges of each feature at baseline and intra-treatment were determined and the change in features was compared using the Wilcoxon signed-rank test.

Of the 29 patients, there were 16 complete responders, 7 partial responders, and 6 non-responders. Comparing CR/PR vs. NR for the gradient-based GTVs, 7 texture values had a $p < 0.05$. The threshold GTVs yielded 4 texture features with $p < 0.05$. ROC and logistic regression was performed and texture features from both PET Edge and thresholding yielding a higher area under the curve (AUC) than SUV metrics. Features derived from PET Edge GTVs also showed higher AUCs than the threshold GTVs. From baseline to intra-treatment, 16 texture features changed with $p < 0.05$. Texture analysis of PET imaged gynecological tumors is considerably more powerful than SUV in early prognosis of tumor response, especially when using a gradient based method.

We then took the 16 texture features showing significant changes ($p < 0.05$) between baseline and intra-treatment PET scans in 29 patients and tested these against the subset of reconstructed features to determine if these changes were dependent upon the method in which the scans were reconstructed. A total of 13 features (including entropy, zone non-uniformity, and complexity) were found to be consistently different

even when subjected to different means of reconstruction, however 3 of the 16 (inverse variance, run percentage, and zone percentage) were found to be dependent upon these reconstruction parameters. Texture features such as entropy, zone non-uniformity, and complexity are excellent candidates for future investigations of changes in texture analysis during radiation therapy of gynecological cancers. Caution should be taken with inverse variance, run percentage, and zone percentage due to their dependence upon reconstruction parameters.

This comprehensive work characterizes gynecological cancers using texture analysis in order to identify texture features that may be used for predicting tumor response as well as reflecting changes during treatment. It is the first study to our knowledge that utilizes all 4 texture matrices (GLCM, GLRLM, GLSZM, and NGLDM) and found 7 statistically significant features classifying responding and non-responding gynecological tumors: *energy*, *entropy*, *max probability*, *zone gray level non uniformity*, *zone size non uniformity*, *contrast (NGLDM)*, and *complexity*. A novel method was implemented extending the NGLDM and its respective features to 3D space for this study. It is also the first study concluding that a semi-automatic gradient-based segmentation method results in more, stronger predictors than using a 40% SUV_{max} threshold method. Finally, this is the first study to examine variability of texture features with reconstruction parameters and to identify texture features as reliable and

independent of reconstruction. In conclusion, texture analysis is a promising method of characterizing tumors visible on PET and should be considered for future studies.

Dedication

This thesis is dedicated to my parents, Ruth Wade and Joe Nawrocki, for always encouraging me to follow my passions, teaching me to love learning, and taking me on the many adventures we have had exploring the world. I could not have done this without your support.

Contents

List of Tables

List of Figures

List of Abbreviations

ADC	Apparent Diffusion Coefficient
AUC	Area Under the Curve
CR	Complete Responder
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
DSC	Dice Similarity Coefficients
FDG	Fluorodeoxyglucose
GLCM	Gray Level Co-Occurrence Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
GTV	Gross Tumor Volume
GUI	Graphical User Interface
HGRE	High Gray Run Emphasis
IRB	Internal Review Board
MR	Magnetic Resonance
MV	Metabolic Volume
NGLDM	Neighborhood Gray Level Difference Matrix
NR	Non-Responder
OSEM	Ordered Subset Expectation Maximization

PET	Positron Emission Tomography
PR	Partial Responder
PTV	Planning Target Volume
ROC	Receiver Operating Characteristics
TLG	Total Lesion Glycolysis
TOF	Time of Flight
ZSNU	Zone Size Non Uniformity
ZGLNU	Zone Gray Level Non Uniformity

Acknowledgements

I would first like to first thank my advisor, Dr. Oana Craciunescu, for guiding me along the path of research. Without her motivation, support, and mentorship, this thesis would have never been accomplished. Additionally, I would like to thank Dr. Junzo Chino for providing the clinical support and background for this study, Dr. Shiva Das for assisting with the early coding, Dr. Tim Turkington for his insight into the technical aspects of PET, and Dr. Irina Vergaslova and Kim Light for helping me learn Eclipse. Dr. Vikren Sarker and Dr. Bill Salter, who were my mentors between undergraduate and graduate school as well as between my first and second year of graduate school, provided valuable assistance with DICOM formatting and navigating RT structure files.

Many students also helped with this project including Chunhao Wang who suggested the 3D correction to the NGLDM, Kyle Higgins for the assistance with statistics, and Kelly Loman who helped with debugging code as well as extending the project to MRI. Finally, I would like to thank the students of the Radiation Physics lab who provided support throughout this work.

1. Introduction

Radiation oncology utilizes medical imaging in all steps of the treatment process, from the first simulation scan used for treatment planning, to daily patient position, to post-treatment follow-up[1]. Additionally, there is an increasing focus on developing medical imaging, in conjunction with genetic biomarkers, as a method for the early prediction of tumor response to radiation therapy. The 2014 keynote address at the annual meeting of the American Society for Radiation Oncology (ASTRO) as well as the 2015 Science Council Session at the annual meeting of the American Association of Physicists in Medicine (AAPM) both stress the use of imaging in deriving physiologically meaningful features in order to gain further insight into how a tumor will respond to therapy. Methods to achieving this goal include bioinformatics and radiomics; both are umbrella terms encompassing a wide range of methods. Bioinformatics includes the statistical analysis of gene maps in order to identify biomarkers linked to cancer, while radiomics is a general term applied to techniques used to derive quantitative features from medical imaging. Texture analysis falls within radiomics and is a broad term applied to techniques utilizing mathematical algorithms classifying an image based on the distribution of pixels.

Originally, texture analysis was designed in the 1970's to automatically classify aerial images into categories of land types[2]. Based on the features derived, tracks of forest, desert, or urban sprawl could be quickly marked for any number of reasons.

Texture analysis was quickly adopted by industry as a method of material inspection[3]. Several decades later, health sciences explored texture analysis as a technique to classify cellular structure[4]. Most recently, texture analysis has been applied to CT, MRI, and ultrasound imaging for automatically segmenting organs[5]. In PET imaging, texture analysis is showing considerable potential for predicting if a tumor will respond to therapy or not[6, 7]. For example, a tumor visible on a given image may contain valuable structural information regarding the aggressive nature of the tumor. Visually discerning structural patterns can be very difficult. Texture analysis of PET imaged tumors can quantify features corresponding with tissue structure based on patterns of pixels and relate these features with tumor response.

This comprehensive thesis examines texture analysis of node positive gynecological malignancies in the context of an adaptive ^{18}F-FDG PET radiation therapy protocol. The purpose of this thesis is to establish texture analysis as a reliable imaging tool to be used complimentary to SUV metrics for predicting and assessing gynecological tumor response to radiation therapy. The structure of this paper is organized into chapters designed to provide the reader with a clear narrative. The introduction provides the necessary technical background relating to gynecological malignancies, PET imaging, and radiation therapy. Chapter 2 explores the advantages of adaptive PET protocols in reducing dose to normal tissues. Chapter 3 confirms that for gynecological malignancies, PET Edge is a reliable segmentation method. Chapter 4

presents the four texture matrices and respective texture features used in this study as well as reviews literature regarding texture analysis of malignancies using PET imaging. Chapter 5 documents the MATLAB graphical user interface (GUI) developed for quick, easy texture analysis of PET and MR image sets. Chapter 6 presents the results regarding the power of texture analysis in predicting tumor response including the advantage of PET Edge over threshold based segmentation, correlates changes in texture feature with changes in SUV during treatment, and reflects on how texture features quantify physiological characteristics of tumors. Chapter 7 studies the variability of texture features with varying reconstruction parameters. Finally, Chapter 8 draws conclusions from this study.

1.1 Gynecological Malignancies

Gynecological cancers occur in the female reproductive organs including the cervix, uterus, vulva, vagina, ovaries, and fallopian tubes. Cervical and uterine cancers are the most commonly occurring gynecological cancers and the American Cancer Society estimates that approximately 67,000 cases will be diagnosed in 2015. Rates of cervical cancers in the U.S. have decreased as screening techniques become common, however, cervical cancer is still the fourth most common cancer worldwide in women with 528,000 cases diagnosed in 2012[8]. Cervical cancer forms in the lower part of the uterus and may grow into the vagina while uterine cancer occurs in the inner lining of the uterus (endometrium). Vulvar cancer is rarer and occurs on the external parts of the

female reproductive anatomy. Primary cancers originating in the vagina are rarer than cervical and uterine cancers as well. Many times tumors from the cervix extend into the vagina and are diagnosed as vaginal cancer[9]. Gynecological cancers frequently metastasize along a predictable pathway progressing to pelvic and para-aortic lymph nodes[10, 11].

1.2 Radiation Therapy and Adaptive PET Protocols

Radiation therapy is the practice of using ionizing radiation to treat various cancers. Gynecological cancers, such as cervical cancer, are routinely treated with radiation therapy in addition to chemotherapy[12]. Advances in technology such as Intensity Modulated Radiation Therapy (IMRT) have improved treatments of gynecological cancers by reducing dose to sensitive tissues such as the bladder, rectum, and bowel[13]. In radiation therapy, treatment plans are developed using CT image sets in which both the normal anatomy and region of disease is contoured. The gross tumor volume (GTV) is defined as the area of the tumor visible via CT, PET, or MR. The GTV is expanded to the clinical target volume (CTV) to include areas of microscopic disease not visible with imaging. Finally, the planning target volume (PTV) is created to include the total area that will be treated with radiation. With gynecological malignancies showing nodal involvement, the PTV generally includes the regional lymph nodes[14].

The objective of an adaptive study is to acquire an additional PET-CT scan for planning after a prescribed dose is delivered, re-contour the volumes, and create a new

treatment plan in order to reduce dose to normal tissue while maintaining the prescribed dose to the PTV. As the gross tumor responds to therapy, the reduction in size may cause a shift in both the PTV and normal tissue. Sensitive tissues such as the rectum, sigmoid, and bowel generally experience high dose because of the proximity to gynecological cancers. If the GTV shrinks during treatment, these tissues may experience an increase in dose if margin between the PTV decreases. Therefore, re-planning the treatment may help account for a decrease in PTV size and spare the normal tissue unnecessary dose. PET imaging in adaptive protocols is particularly advantageous because in addition to accounting for possible changes in anatomy, changes in tumor metabolic activity can be assessed.

1.3 PET Imaging

The advent of PET imaging quickly became a standard technique used in quantitatively deriving the metabolic activity of cancers, such as cervical tumors[15]. A glucose molecule tagged with radioactive fluorine, ^{18}F-FDG (fluorodeoxyglucose), injected into a patient will act as a glucose analog and concentrate in metabolically active regions, such as cancer cells. ^{18}F is a positron emitter with a half-life of approximately 110 minutes. As ^{18}F decays, the ejected positron interacts with an electron after losing its kinetic energy and annihilates, resulting in a pair of photons. In order to conserve momentum, the photon velocities are 180^{o} apart. Image reconstruction depends on the coincidence measurements of these two photons. If two photons are detected directly

opposing each other within a small time window, they are considered to have originated from a single ^{18}F atom. Modern PET scanners then use an iterative reconstruction method, such as ordered subset expectation maximization (OSEM) to produce a coherent image set. The number of iterations and subsets used in reconstruction impacts the image resolution and quantitative results and will be further investigated in this work. Improvements to PET reconstruction include a time-of-flight (TOF) correction which improves the signal-to-noise ratio by approximating the location of the decaying atom inside the body by the difference in detected times of the two photons[16].

A combined PET-CT scanner used in many radiation oncology clinics allows for automatic registration of the PET and CT image sets, thus providing both metabolic and anatomic information. The high signal from the PET positive tumor and nodes in contrast to low signal in the rest of the body allows the clinician to more easily contour the target volumes. The PET positive target contours are then transferred to the CT image set to be used in treatment planning.

PET imaging provides the clinician with quantitative values representing the metabolic activity of the tumor in the form of Standard Uptake Volume (SUV). SUV (g/ml) is a near unit-less metric derived from the injected activity of the radioactive tracer as well as the mass of the patient. Accounting for patient body mass as well as the volume and activity of the tracer should, in theory, normalize SUV between patients and

scanners. Different reconstruction methods and scanner calibrations may result in differences in SUV metrics however. Equation 1 presents the calculation of SUV.

$$SUV = \frac{Radioactivity\ Concentration\ (Bq/ml)}{Injected\ Dose(Bq)/Body\ Mass\ (g)} \quad (1)$$

SUV has become a valuable tool in radiation oncology as it tends to correlate with changes in the tumor during radiation[17-21]. The increase or decrease of SUV from baseline PET-CT scans to scans acquired during therapy have been shown to predict if the tumor will respond to treatment[17]. One study found that a decrease of 60% in SUV_{max} from the baseline PET-CT and to a PET-CT during week 4 of concurrent chemoradiation therapy for cervical cancer was a significant predictor of complete tumor response[18]. While studies such as the previous example have examined how changes in SUV during treatment correlate with treatment outcome, only a handful of studies have concluded that baseline SUV alone may show potential in predicting tumor response[19]. Metabolic volume (MV) defined as the total volume in ml of the primary tumor and Total Lesion Glycolysis ($TLG = MV \times SUV_{mean}$) have also been investigated as possible physiological features correlating with tumor response[20, 21]. The primary shortcoming of image-based tumor response assessment, such as SUV, is that it takes both the baseline and intra-treatment image sets to determine if the tumor may respond to treatment, leaving little time to adjust therapy parameters. Predicting tumor response from baseline scans alone would present a clear advantage, allowing for treatments to be adjusted from the beginning.

2. Evaluation of Target and Normal Tissue Doses in an Adaptive PET Protocol for Gynecological Cancers

This study presented a prospective, IRB approved protocol enrolling a subset of 16 women with PET visible malignancies: 10 with cervical cancer, 4 with uterine cancer, and 2 with vulvar cancer. A baseline PET-CT was acquired for treatment planning and normal tissues were contoured. The GTV was contoured by an experienced physician. Prescribed dose varied between 45-50.4 Gy, with a 55-70 Gy boost to the PET positive nodes. An intra-treatment PET-CT was acquired between 30 and 36 Gy. All target and normal tissues were contoured and an adapted plan was developed. The two CT sets were registered using rigid registration on bony landmarks and adjusted based on the high dose PTV location. In order to assess how much dose would have been delivered to normal and target volumes if there was no adaptive plan, an unadapted plan was created by applying the baseline treatment plan to the intra-treatment PET-CT. Doses to 2cc (D2cc) of the bladder, bowel, rectum, sigmoid, and PTV were acquired for both the adapted and unadapted plans and were compared using the Wilcoxon signed-rank test.

As Table 1 presents, dose to the PTV for both adapted and unadapted plans remained nearly identical while dose to the normal tissue volumes was significantly lower in the adapted than unadapted plan. Both the bowel and rectum benefitted most from the adaptive plan ($p < 0.0001$).

Table 1: Median Dose (IQR) to normal tissue and boost PTV with PET adapted plans and unadapted plans

	Bowel D2cc	Rectal D2cc	Sigmoid D2cc	Bladder D2cc	boost PTV D95
Unadapted	55.4 (51.3-59.4)	50.2 (49.2-54.3)	54.4 (50.2-56.8)	50.2 (48.3-54.5)	62.5 (60.7-64.2)
Adapted	52.6 (50.5-53.5)	49.4 (47.9-53.3)	52.8 (49.4-53.6)	49.8 (48.1-52.7)	62.3 (60.5-63.7)
***P* value**	< .0001	< .0001	.0006	.0007	.17

Adaptive protocols for gynecological malignancies present an opportunity to ensure the dose to normal tissue is kept within the limits while maintaining the prescription dose. This study will be continued in the future as more patients are added to the protocol.

3. PET Edge Segmentation Comparison

Several segmentation methods currently exist for delineating PET positive target volumes. Automatically thresholding based on 40% of the maximum SUV in the target is a common clinical and research method used for both SUV and texture analysis. Other methods used primarily in research include fuzzy locally adaptive Bayesian analyses utilizing image noise to automatically segment volumes[22]. Gradient-based algorithms use intensity edge-detection to match, if not improve, on thresholding techniques for segmenting PET positive volumes[23-25].

PET Edge is a semi-automatic gradient-based algorithm implemented in MIM 6.0 for semi-automatic segmentation and is evaluated in this study. The primary advantage for using this method is the ease-of-use and reproducibility. It is important to keep in mind that like any semi-automatic method, the user must review the contoured structure to ensure it is adequately delineated. Previous studies have already shown that PET Edge can reliably and reproducibly segment volumes when compared to the threshold method. One study found that the gradient-based method matched maximum diameters of surgical removed lesions better than threshold segmentation[26]. In this study, PET Edge is compared to the physician contours in order to justify the use of PET Edge in further, in-house research projects.

3.1 Comparison between PET Edge and Manual Segmentation

Thirteen women with PET positive gynecological tumors were considered for this study, 11 of which had PET visible nodal involvement. The primary and nodal PET GTVs of both baseline and intra-treatment scans were first contoured by an experienced physician. The same volumes were recontoured using PET Edge in order to compare the two methods. Primary and nodal GTV volumes were measured and compared using Bland-Altman plots in order to determine the statistical significance of the difference between segmentation methods. Additionally, the Dice Similarity Coefficient (DSC) was determined for each GTV comparison. DSC is a measurement of spatial overlap between two areas and has been previously used to compare and validate segmentation methods in radiation therapy[27, 28]. Consider two volumes: volume A and volume B. The measured overlap is determined by the ratio of the shared volume to the combined, independent volumes, presented as

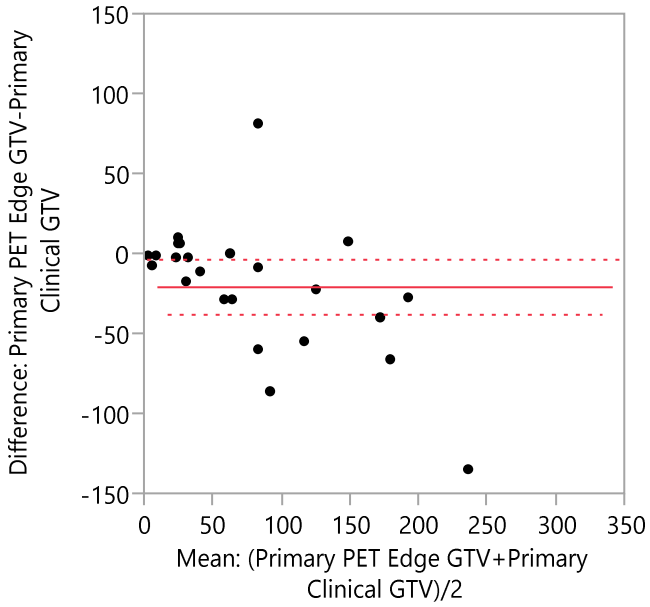
$$DSC = \frac{2|A \cap B|}{A + B} \tag{2}$$

A perfect match would be equal to 1. Using the Boolean operators within MIM, the physician and PET Edge contours per image set for both primary and node GTVs were combined into a single contour, representing A ∩ B. DSC values for thirteen primary GTVs and eleven node GTVs were calculated.

The Bland-Altman plots presented show the statistical distribution of clinically contoured volumes relative to PET Edge volumes. The horizontal axis presents the

mean values of corresponding primary and node GTVs, in increasing volume (cc). Percent deviations of the clinical contours from the mean are plotted in the vertical axis. Values within 1.96 σ (95% confidence interval) are generally considered to be clinically acceptable.

(a)



(b)

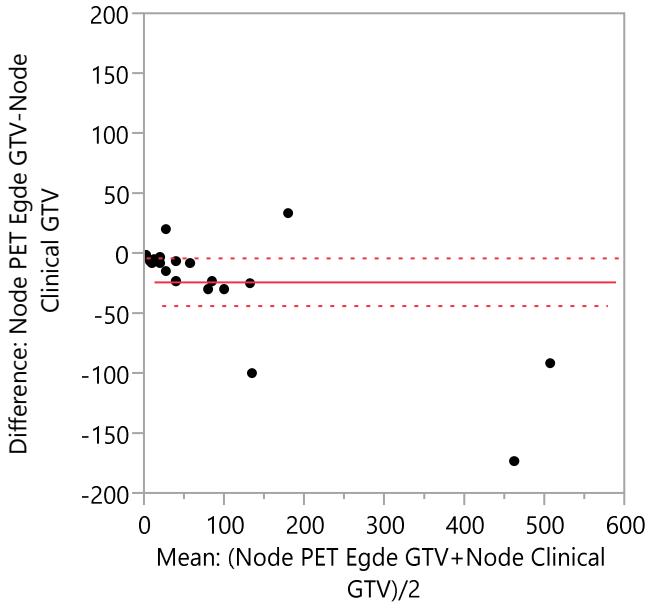


Figure 1: Bland Altman plots depicting differences between the clinically and PET Edge contoured (a) primary GTV and (b) node GTV volumes (cc).

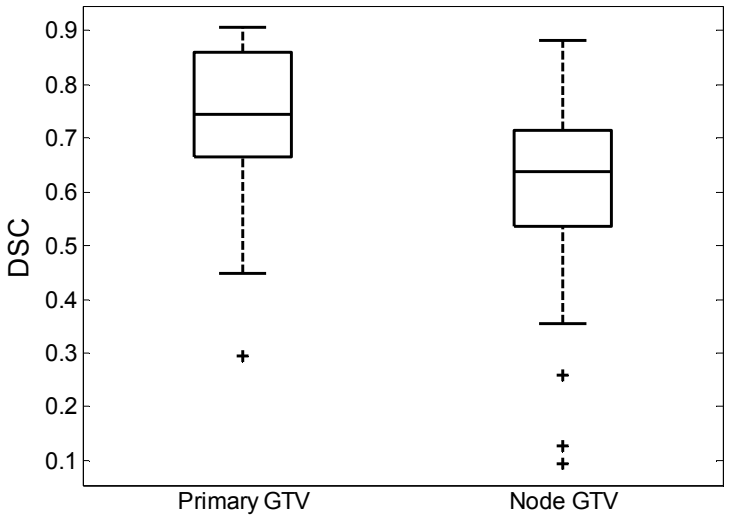


Figure 2: Dice Similarity Coefficients for comparing clinically and PET Edge contoured primary and node GTVs.

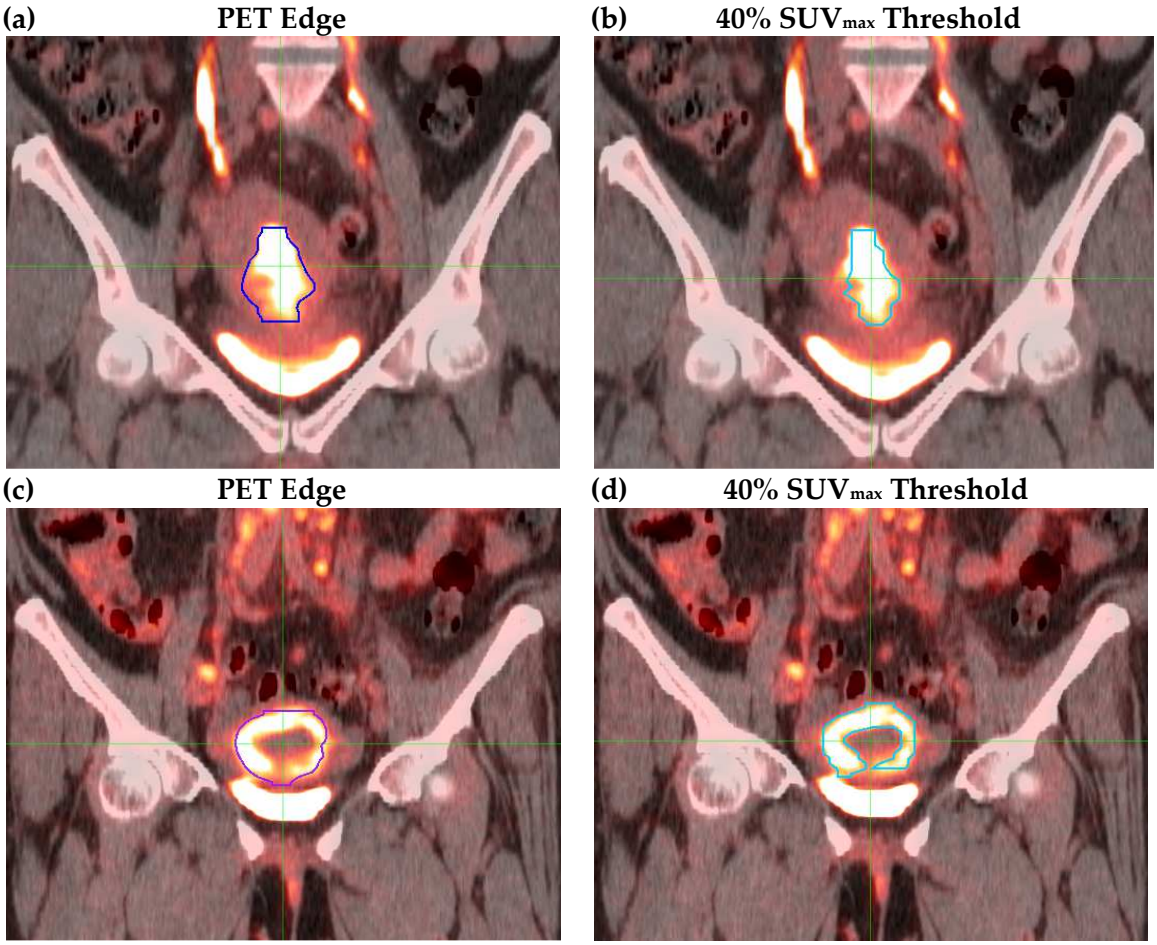


Figure 3: Comparison of PET Edge (a and c) and threshold (b and d) contoured GTVs. The two images on top represent a scenario in which PET Edge and thresholding yields very similar contours while the two images on bottom represent a scenario in which the two contours are fairly different.

Figure 1 shows that the PET Edge contoured GTVs tend to be smaller than the clinically contoured GTVs. Small volumes for both primary and node GTVs tend to group closer to the mean value. Considering the size of the smaller GTVs, deviations in the PET Edge contouring will not cause large changes regarding volume. For larger volumes, the clinical GTVs tend to be larger than the PET Edge contours. The two nodal volumes around 500 ml were exceptionally large vulvar case nodes. Deviations in

contouring large volumes will result in large volume differences that do not necessarily indicate proportional error. The DSC values plotted in Figure 2 indicate that the primary GTV clinical and PET Edge contours are reasonably close while the node GTVs are more variable. PET Edge works well for large volumes but as the size decreases, the tool becomes less effective. Figure 3 depicts two scenarios in which PET Edge contours may be similar or different to clinical, thresholded contours. Figure 3.c and Figure 3.d illustrate how PET Edge will generally contours the entire GTV while thresholding will not contour the area of low FDG uptake in the center of the tumor. This difference may account for differences between volumes and DSC coefficients. Overall, PET Edge is an effective tool for contouring PET positive primary GTVs, though it should be checked against the physician contoured GTVs to ensure consistency.

4. Texture Analysis: Theory and Practice

Texture analysis is a method of extracting spatial information regarding pixel intensity relationships in an image. Quantifiable features derived through texture analysis represent underlying patterns within the image which may reflect either the chaotic, or uniform, nature the pixel distribution. The derivation of texture features is accomplished by calculating matrices tabulating visual patterns and applying statistical methods to quantify each pattern. Textures are somewhat ambiguous, and assigning a physically meaningful description to the derived metrics is particularly challenging. A simplistic view is that the texture of an image can either be described as homogeneous or heterogeneous. Homogeneous textures are visually smooth while heterogeneous textures appear random and splotchy.

Quantifiable image characteristics derived from tabulations of pixel relationships can be classified as first, second, and high order features depending on the number of pixels per group. First order features examine groupings of single pixels and are derived from the gray level histogram. While these features describe the image, they do not describe patterns within the image and are therefore not classified as textures. First order features are often referenced as global features because they characterize the entire distribution of data, such as histogram features. A gray level histogram is a distribution of the number of pixels per gray level. In addition to deriving the max, median, and mean pixel values, three features describing the histogram are calculated: skewness,

kurtosis, and variance. Skewness describes the asymmetry of the histogram with zero being a perfectly symmetrical distribution. Kurtosis describes the difference between the peak histogram values and the rest of the data. Variance describes the spread of data. The formulations are presented in Appendix A, Table 11.

Second order features are statistical formulations describing groupings of two adjacent pixels. Features derived from the gray level co-occurrence matrix (GLCM) tabulate pixel pairs and are often referred to as local features as they describe a set number of pixels. High order features examine groups of pixels in regions of the image of variable size. The gray level run length matrix (GLRLM) and the gray level size zone matrix (GLSZM) are high order matrices from which regional features are derived. The gray level neighborhood difference matrix (GLNDM) calculates high order features classified as local because the matrix tabulates a set area of pixel neighborhoods within the image.

4.1 The Gray Level Co-Occurrence Matrix and Local Features

The defining paper presenting the 2D GLCM was authored by Robert Haralick in 1973, giving rise to the frequently cited 'Haralick's texture features'[29]. It was created as a tool to categorize images using computers. The GLCM model developed in this work relies on this paper and extends it to 3D space. Haralick proposed a methodology to generate a matrix with dimensions equivalent to the number of gray level intensity values in the image. The matrix tabulates the number of occurrences between every

pixel in the image. Eleven features are then calculated from this matrix and are designed to quantify visual features that humans understand such as coarseness or smoothness.

Consider an image with four levels of pixel intensity. The corresponding GLCM will be a 4x4 matrix wherein the number of occasions a given pixel pair occur are tabulated.

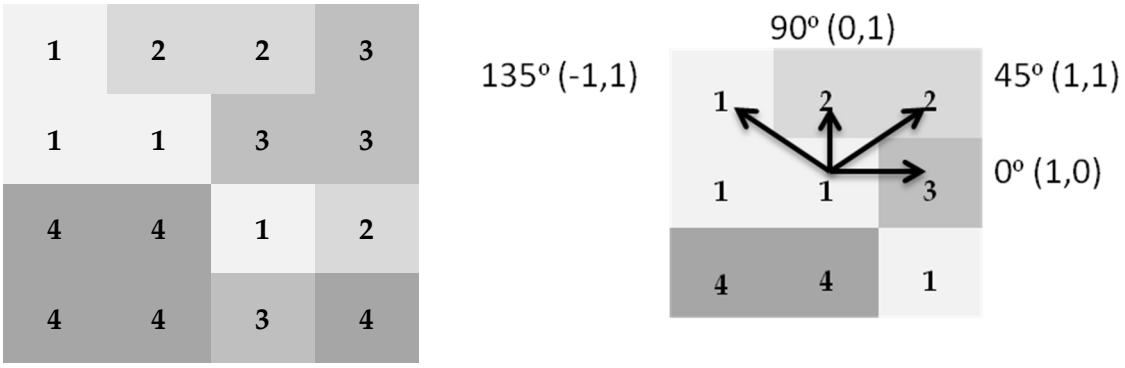


Figure 4: (left) A sample image with 4 gray levels and (right) the 4 2D directions of the pixel relationships from the measurement pixel.

0°	1	2	3	4
1	2	2	1	1
2	2	2	1	0
3	1	1	2	2
4	1	0	2	4

0°	1	2	3	4
1	.083	.083	.042	.042
2	.083	.083	.042	0
3	.042	.042	.083	.083
4	.042	0	.083	.17

Figure 5: (left) The symmetrical gray level co-occurrence matrix tabulating pixel pairs in the 0° direction. (right) The matrix is then normalized by the sum of the contents.

The given example presents the GLCM derived from the example image in the 0^{o} direction. Both the X and Y axis represent the gray level of one of the pixels in the pair. For example, pixels of intensity 1 and 2 exist adjacent to each other, horizontally, on two occasions. In a 2D image, there are four directions available from the reference pixel in the center, covering the entire image space. An additional consideration is the pixel distance used in establishing pixel pairs. In this example, as well as the GLCMs implemented in this work, a pixel distance of one is used. Note that the GLCM is symmetric in order to account for pixel relationships in the opposite directions. After 4 GLCMs are derived, one for each direction, the matrix is normalized by the sum of the contents. This step allows for easier interpretation of the features derived from the GLCM.

Haralick's paper details the GLCM methodology for 2D images, however, tomographic medical imaging includes multiple slices comprising a 3D volume. Co-occurrence matrices can be created in 3D by extending the directions from 4 to 13.

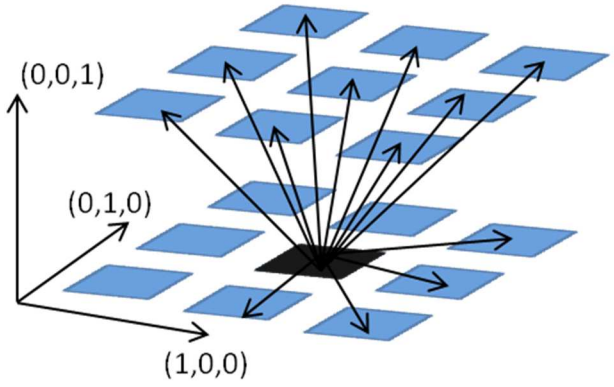


Figure 6: Thirteen directions of pixel pairs to account for a volumetric GLCM.

Figure 6 presents the additional nine directions considered in the 3D model. Symmetry ensures the entire volume is accounted for, totaling 26 directions. Therefore, 13 normalized GLCMs are produced for a single volume.

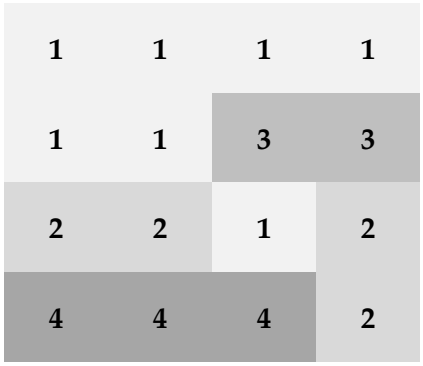
Haralick presents 14 local, 2nd order texture features extracted from each directional GLCM. Only unique features were considered. For example, three features relate to the image *entropy: entropy, sum entropy,* and *difference entropy*. Because both *sum entropy* and *difference entropy* are proportional to *entropy*, and the most commonly referenced feature in the literature is *entropy*, only the strict *entropy* feature was considered. Subsequent papers have suggested additional features derived from the GLCM and were also considered[30]. Appendix A, Table 7 presents the equations used for 11 local features extracted from the GLCM in this study.

4.2 The Gray Level Run Length Matrix and Regional Features

The gray level run length matrix (GLRLM) presented by Galloway in 1974 is an additional texture matrix similar to the GLCM, but represents lengths of pixels instead of pairs[31]. Each gray level run consists of a series of pixels of uniform intensity directed along a linear path. A length may be as short as a single pixel, or as long as the image allows. As in the GLCM, run lengths are calculated throughout the volume according to the thirteen angles previously designated. Because runs of any length are considered, symmetry is inherent to the matrix. In general, fine textures tend to have many small runs of similar pixel intensities while coarse textures contain longer runs of

varying intensities. GLRLM features emphasize regional information instead of the highly localized information of the GLCM. From the GLRLM, we are able to determine if the image contains long or short lengths of high or low intensity gray levels.

The formation of the GLRLM is similar to the GLCM in that the matrix depends on the number of gray levels present designated along the vertical axis. The other component is the length of each gray level intensity run designated along the horizontal axis. For each intensity level along the vertical axis, the number of occurrences of a given length of runs is tabulated in the matrix.



0^o	1	2	3	4
1	1	1	0	1
2	2	1	0	0
3	0	1	0	0
4	0	0	1	0

Figure 7: (left) A sample image containing 4 gray levels and (right) the corresponding run length matrix in the 0^o direction.

Pixels of intensity 1 occur once in a length of 4 pixels, once in a length of 2 pixels, and once in which there is a single pixel present. This tabulation is performed in all directions, producing 13 matrices, as in the GLCM.

Eleven regional, high order texture features were derived from the GLRLM describing features of the image[32]. Features describe the prevalence of shorter lengths,

representing fine structure, or longer lengths, correlating with coarseness. Other features describe the distribution of pixels based on intensity, or various combinations of lengths and pixel intensities. Appendix A, Table 8 presents the formulations used in this work.

4.3 The Gray Level Size Zone Matrix and Regional Features

The GLSZM is a texture algorithm similar to the GLRLM. Instead of calculating lengths of pixels however, the GLSZM calculates zones of uniform pixels in an image[33]. In a 2D image, a zone would consist of any adjacent pixels of common intensity in the image. This model is extended to a volume by finding groups of uniform pixels in each of the 26 available directions in 3D space. The matrix tabulates the occurrences of gray levels grouped in zones of any size. As with the GLRLM, the GLSZM characterizes regional texture features. Because the zones are directionally independent however, only a single matrix is produced.

1	**1**	**3**	**1**
1	**1**	**3**	**3**
2	**2**	**3**	**2**
4	**4**	**2**	**2**

	1	**2**	**3**	**4**
1	1	0	0	1
2	0	1	1	0
3	0	0	0	1
4	0	1	0	0

Figure 8: (left) A sample image containing 4 gray levels and (right) the corresponding size zone matrix.

The GLSZM is similar to the GLRLM in both design and purpose. All 11 regional, high order features derived from the GLRLM are applied to the GLSZM with the zones substituted for the run lengths and are listed in Appendix A, Table 9. Analyzing the distributions of uniform volumes of pixels within an image is perhaps more intuitive than extracting lengths from the volumes. A homogenous image may have a more uniform distribution of zones while a heterogeneous image may have a high variation in zones.

4.4 The Neighborhood Gray Level Difference Matrix and Local Features

Previous algorithms describe pairs or groupings of pixels and extract features reflecting characteristics resembling what our own visual conclusions may be. The NGLDM presents a more complex set of features. Instead of grouping pixels based on intensity, the NGLDM takes an average of the surrounding pixel neighbors and in order to determine the variation in texture[34]. In other words, this texture analysis approach examines changes in intensity between the target pixel and the surrounding neighbors. The relationship between the averaged neighborhood values and the target pixel, such as small or large differences, reflect the texture of an image.

Originally, the NGLDM was proposed for 2D images. The functional form of the NGLDM is presented as

$$s(i) = \sum |i - \bar{A}_i| \tag{3}$$

where

$$\bar{A}_i = \frac{1}{W-1}\left[\sum_{m=-d}^{d}\sum_{n=-d}^{d} f(k+m, l+n)\right] \tag{4}$$

$$(m,n) \neq (k,l)$$

W is a normalization factor to account for the neighborhood size.

$$W = (2d+1)^2 \tag{5}$$

d represents the distance away from the center pixel to the neighborhood pixels. In this work, d = 1. For each intensity level, the surrounding neighbors to the respective center pixel of those intensities are added, normalized, and then compounded. Figure 9 better illustrates a 2D neighborhood in an image.

1	**1**	**1**	**1**
1	**3**	**3**	**4**
2	**2**	**1**	**2**
4	**4**	**4**	**2**

Figure 9: A sample image with 4 gray levels. The box indicates the neighborhood surrounding a central pixel.

In the designated area of this image, the center pixel has a value of 3. The total value of the neighbors is 12. For this pixel, $\bar{A}_i = \frac{12}{8}$. There are two pixels of value 3 in this image however. The NGLDM entry $s(3) = \left|3 - \frac{12}{8}\right| + \left|3 - \frac{15}{8}\right| = 2.625$. This is

performed for every intensity level. The length of the matrix depends on the number of gray levels in the image.

In this work, the NGLDM is extended to 3D. Extending this matrix to 3D is not presented in the literature, however, it is a fairly straightforward process following the logic of the algorithm. In the calculation of the matrix, the number of available neighbors is 26 instead of 8, accounting for the area of a cube surrounding the center pixel. The 3D weighting factor becomes $W = (2d + 1)^3$. Since the calculation is directionally independent, only one matrix is produced for a 3D volume.

The features presented in the original paper were created for a 2D image. Certain changes must be applied to account for 3D space, most of which substitute the 2D weighting for a 3D weighting factor. There are 5 local, high order features derived from the NGLDM. Formulations presented in Appendix A, Table 10 are the corrected 3D equations for the NGLDM features. It is encouraged to reference the original paper in order to clarify volumetric corrections.

4.5 Implementation of Texture Analyses for PET Images

The algorithms presented describing texture matrices must be applied to the PET GTV. Each pixel in a PET image represents the SUV for a 3D voxel. Texture analysis requires discrete values in order to tabulate the gray levels into matrices. The pixels in the GTV must be re-sampled to a normalized value between patient image sets.

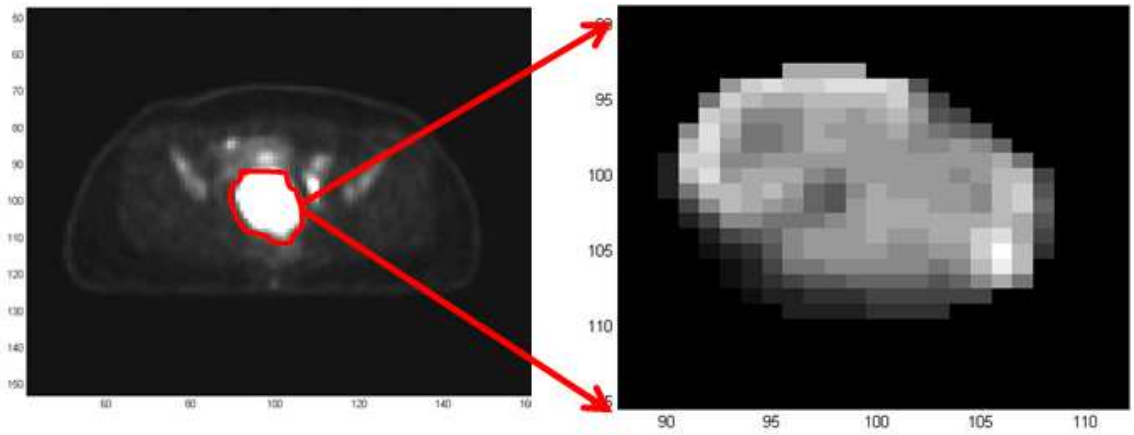


Figure 10: The GTV is extracted from the original image and normalized to 64 bins.

Normalization allows for tumor textures to be compared between patients with varying uptake levels and minimizes the effects of noise[38]. The method used in this work is defined as:

$$I_{Norm} = \frac{2^n\left(SUV_{pixel} - SUV_{min}\right)}{SUV_{max} - SUV_{min}} \tag{6}$$

Various re-sampling levels may be used, as will be explained later, however, re-sampling to 64 (n = 6) is used in this study as this value has been found in the literature to yield texture features the correlate best with SUV metrics[40]. The four texture matrices are applied to the normalized GTV with only the GTV being considered for. Any pixel of value 0 is not considered in the algorithms.

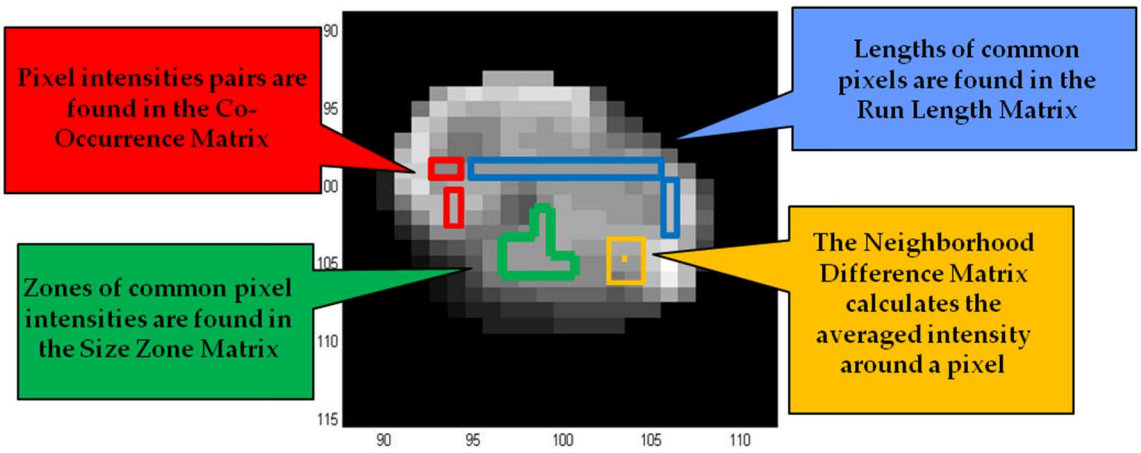


Figure 11: Example from each texture matrix overlaid on a sample slice.

4.6 Texture Analysis for Tumor Response and Assessment using PET

Having defined the four texture matrices, it is important to review the current literature on texture analysis in PET imaging. Various studies explore texture analysis of PET images in multiple sites such as lung, esophageal, and cervix for predicting tumor response and tumor changes during radiation therapy[35-37, 40-42]. Though not all papers reach the same conclusions, features such as entropy and energy generally appear as a statistically significant predictor of tumor response[35]. *Homogeneity* and *high gray run emphasis* (*HGRE*) are identified as significant predictors of breast tumor response in another study [36]. *Complexity* has also been suggested as a potentially significant texture feature in breast cancer[37].

Tixier et al. presents a study in which texture analysis was used for predicting tumor response of 41 patients with esophageal cancer. PET primary tumors were

delineated using a fuzzy locally adaptive Bayesian algorithm and 31 texture features were derived from the 4 texture matrices in addition to SUV histogram features. Texture features highlighted as predictors of responding and non-responding tumors include *entropy, homogeneity, intensity variability* (same as *ZGLNU*), *size zone variability* (same as *ZSNU*), and *coarseness*. Tumor response was assessed on post-treatment CT based on Response Evaluation Criteria in Solid Tumors (RECIST) [38].

Table 2: Statistically significant features predicting tumor response found by Tixier et al. SUVmax added for comparison

Feature	p	AUC
Homogeneity	> 0.05	0.89
Entropy	0.0006	> 0.82
Intensity Variability	0.0002	> 0.82
Size Zone Variability	0.0002	> 0.82
Coarseness	0.0002	> 0.82
SUV_{max}	0.106	0.59

Another highly sited paper on texture analysis of PET tumor response by Orlhac et al. including 28 patients with metastatic colorectal cancer, 24 patients with non-small cell lung cancer, and 54 patients with breast cancer. Primary lesions were delineated using two methods: VOI_{TAB} (modified version of thresholding based on 70% maximum

intensity described by Nestle et al.) and 40% SUV_{max} thresholding[39]. In addition to SUV histogram metrics, 31 texture features from the 4 texture matrices were derived and correlations between SUV and MV determined. They conclude that texture features show weak correlation with SUV, but strong correlation with MV. Dependence on segmentation method was examined and found that many of the GLRLM and NGLDM features were highly dependent on segmentation method while most other features showed little variability. Their study determined that *homogeneity*, *entropy*, *short run emphasis* (*SRE*) and *zone size non-uniformity* (*ZSNU*) are strong candidates for further studies into tumor response because of their independence from segmentation method. Further, these features do not directly correlate with each other, signifying that they characterize different textures within the tumor[40]. The reader should refer to this paper as Orlhac et al. presents a clear summary of several site specific papers indicating texture features found to be effective predictors, correlation, resampling schemes, and segmentation methods.

Two papers including texture analysis of cervical tumors using PET imaging include Naqa et al. and Yang et al. Published in 2009, Naqa et al. presents one of the earlier papers using texture analysis in PET imaging and focus on 4 features from the GLCM: *energy*, *contrast*, *homogeneity*, and *entropy*, plus SUV features (maximum, minimum, and mean) and MV. Primary tumors of 14 cervical patients were segmented using a 40% threshold method. ROC analysis was performed for MV, SUV, and texture

features. *Energy* presented an AUC of 0.725 while the other texture features all had AUC < 0.7. SUV metrics showed AUCs of 0.459 and MV had an AUC of 0.571[41].

Yang et al. utilized the GLRLM and GLSZM in predicting and assessing cervical tumor response to chemotherapy and radiation therapy. Primary PET tumors were segmented via 40% threshold and 11 features from each texture matrix were calculated. Their results show nearly all texture features and SUV metrics showed a statistically significant change from the baseline PET-CT to a PET-CT acquired after 4 weeks of treatment. No texture features from the GLRLM and GLSZM or SUV metrics derived from the baseline PET were able to statistically classify tumors into responders or non-responders[42].

Inconsistencies between papers include varying normalization schemes and texture feature definitions. The method of normalization may lead to varying outcomes however. Tixier et al. suggests using the following:

$$I_{Norm} = \frac{2^n\left(SUV_{pixel} - SUV_{min}\right)}{SUV_{max} - SUV_{min} + 1} \tag{7}$$

Orlhac et al. argues for using Equation 6. Both Equation 6 and Equation 7 were tested by Orlhac et al. Of the 31 texture features derived, 19 were found to have little dependence, 10 were found to moderate dependence, and 2 were found to have large dependence on the re-sampling scheme. Equation 6 yielded texture feature showing a stronger correlation with SUV metrics, and is therefore adopted in this work. The normalization level used varies from 4 to 128 bins depending on the study, each with

differing results. Normalization does affect textures as fewer bins will increase the homogeneity of the image. As previously mentioned, this work uses 64 bins. Orlhac et al. compares texture analysis using 8, 16, 32, 64, and 128 bins and found that the results correlated best with normalization level greater than 16 bins.

The formulation for the metric *homogeneity* changes between papers as well. It should be noted that the original version as it appears in Haralick's paper is presented as:

$$Homogeneity = \frac{\sum_i \sum_j P(i,j)}{1 + (i-j)^2} \tag{8}$$

The alternatively presented formulation (Equation 9) was tested in this work to determine if there was a significant impact on correlation:

$$Homogeneity = \frac{\sum_i \sum_j P(i,j)}{1 + |i-j|} \tag{9}$$

Consistency in methodology is critical when comparing studies of texture analysis. While most studies generally follow a similar process, it is important to consider what differences in methodology may yield varying conclusions such as differences in contouring or defined equations.

5. Duke FIRE: Functional Imaging Research Environment

In order to implement texture analysis in a clinical environment, it is necessary to program a user friendly interface with a minimal learning curve. Using MATLAB, a Graphical User Interface (GUI) was developed to import axial images and structure sets in DICOM format from any other clinical platform, select the GTV, and export an excel file containing the derived features. The GUI, entitled Duke FIRE, was then produced as an executable file and installed on a computer in Radiation Oncology. PET DICOM files can be easily exported from MIM to the clinical computer and imported into Duke FIRE. Duke FIRE can also currently analyze axial MR apparent diffusion coefficient (ADC) images and will soon be extended to coronal MR images mapping several pharmacokinetics parameters. Duke FIRE was designed to have an intuitive interface (see Figure 12) and simple data exportation, allowing for quick use by dosimetrists, physicians, and physicists.

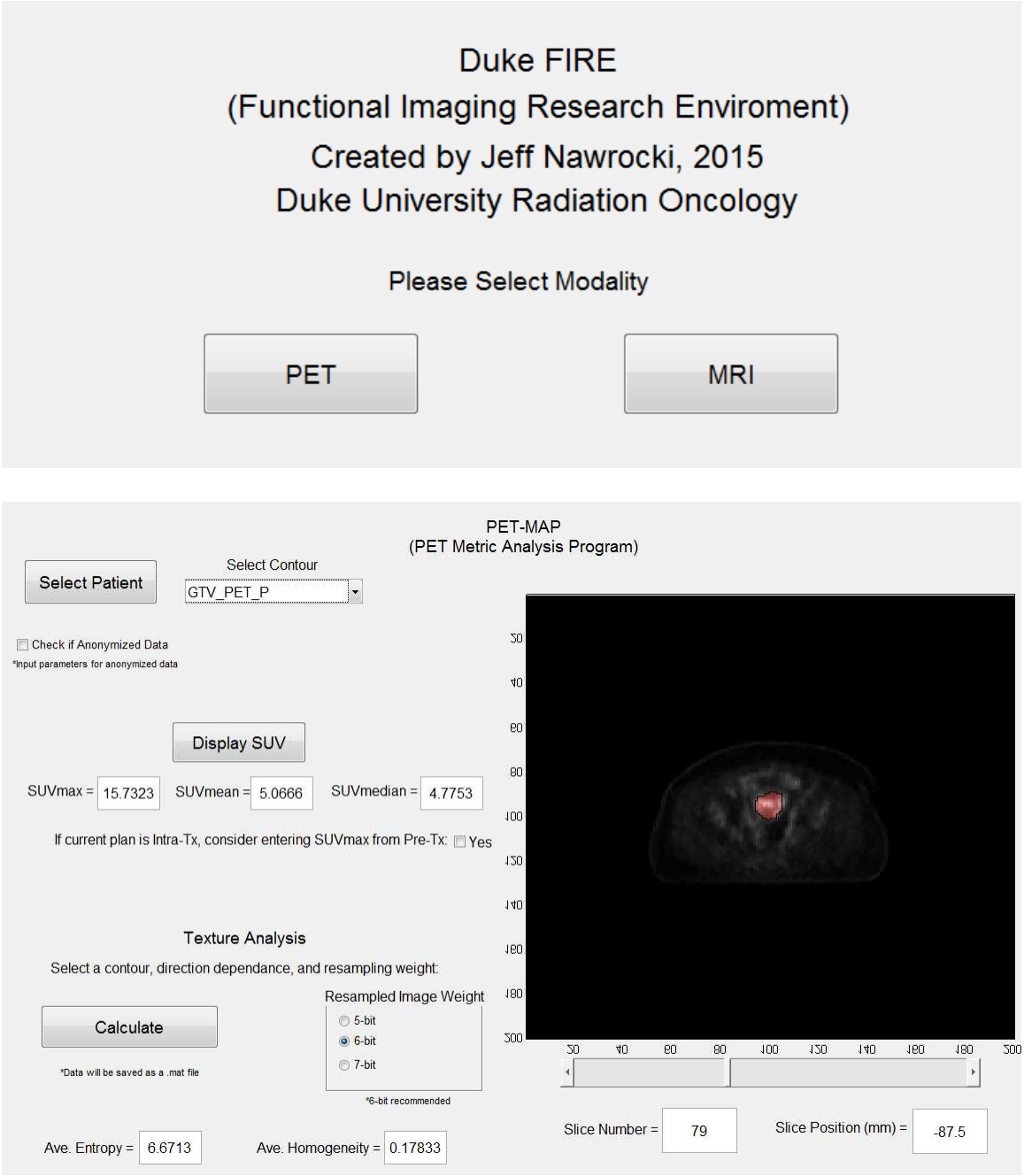


Figure 12: Duke FIRE offers intuitive texture analysis for both PET and MRI. After selecting the modality, the user is able to import an image and contour set, view the images and contours, calculate texture and functional metrics, and export a spreadsheet.

Figure 13 presents a workflow documenting the steps taken in Duke FIRE. The user selects the folder to import DICOM files from and chooses the appropriate GTV from the 'Select Contour' dropdown menu. After calculating the features, the user saves the data to an excel spreadsheet. Note that this program was developed using several functions for exporting excel files unique to the Windows operating system.

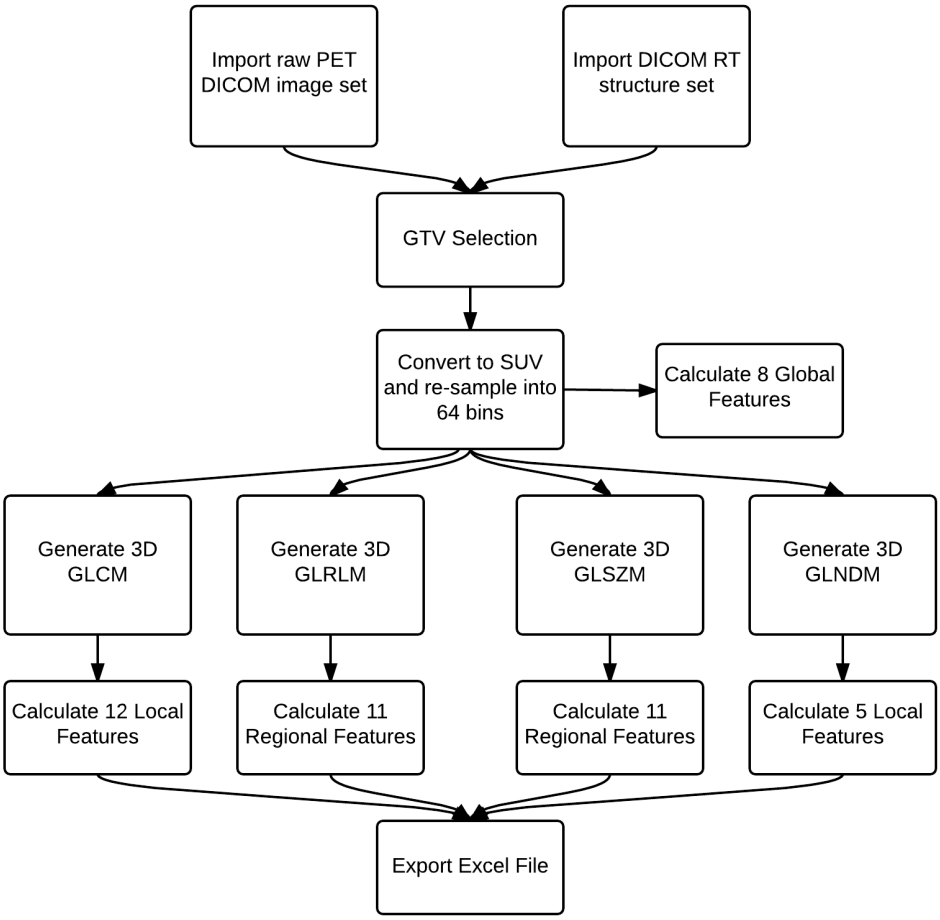


Figure 13: Duke FIRE follows an import-export workflow in which the calculations are automatic. This includes an image sorting algorithm, adaptive SUV calculation, and structure-based data storage.

Developing the GUI required several key aspects: understanding and manipulating the DICOM format, contour selection, SUV calculation, and finally implementation of texture algorithms.

5.1 Brief Summary of DICOM

NEMA (National Electrical Manufacturers Association) and the ACR (American Board of Radiology) created DICOM (Digital Image and Communications in Medicine) as a standard for medical image data. CT, MR, and PET image sets as well as structure sets used in radiation oncology are saved in the DICOM (.dcm) format in order to create a uniform data structure across different platforms. While the principle behind DICOM is to create a single image and communication platform consist across the board, DICOM formatting frequently changes between manufacturers, image viewing, and treatment planning software. Descriptions of DICOM presented in this work therefore pertain to the specific scanners and planning software used herein and may not apply to other platforms. The PET images were produced using a Siemens Biograph PET-CT. The planning software from which the PET images were exported was either MIM or Velocity.

DICOM includes two primary sections: image and header information. Image information contains the array of data comprising the image, typically compressed images in a 16-bit format. For CT and MR images with data ranges of ranging upwards of 2000, a 16-bit range is more than enough. PET image data however, is a measurement

of activity (Bq/ml) and therefore contains a much larger range of data. The images are compressed and a scale factor is created which, when multiplied across the compressed image data, produces images in units of Bq/ml. The scale factor (RescaleSlope) is found in the header information produced along with the images.

Header information contains all information associated with the images including patient names, dates of birth, physicians, and scan dates and times. PET DICOM files contain information regarding the injected radioactivity of FDG as well as injection time, both of which are critical in determining SUV. Each image has corresponding header information, including a RescaleSlope value mentioned above. It is important to note that this scaling value may change per individual image.

Header information for radiation therapy structure sets contains the XYZ locations of the volume contours as well as pixel dimensions in units of mm. Patient position, such as head-first supine, is included as well which plays an important role in where DICOM references the image (0, 0) point. RT structure information is filed in DICOM according to each contoured volume. A segmented volume will consist of multiple slices, and may contain several independent sections in a single axial view. Every independent section per axial slice is designated a separate location in DICOM. For example, a lung contour on a single axial slice will present to sections, one for each lung. In DICOM, the contour XYZ data are filed separately with only the Z (inferior-superior) information remaining constant. It is the Z information that allows for

determining which contour sections belong to a single slice. MATLAB utilizes two functions in reading a DICOM formatted object: *dicominfo* and *dicomread*. *dicominfo* interprets the header information and produces a structure file with the corresponding data. *dicomread* reads the image data and produces the corresponding array.

5.2 Image Import and Contour Selection

The first function of the GUI is importing the DICOM image and structure set of a given patient. After selecting the folder containing the DICOM files, the code reads the header information. If the modality detected is 'PT', it is a PET image. The images are sorted based on the slice location and each image's header information is recorded. If the modality detected is 'RTSTRUCT', the structure set is recorded and the volumes are produced in a list. After the images are sorted and printed to the GUI, the user may select the GTV, or other volume, from the menu of contoured volumes. The selected volume will appear highlighted on the PET image.

In order to process the GTV and calculate texture features, the GTV image data must be segmented, or isolated, from the rest of the image. This is performed by reading the selected contour's XYZ data from the structure set. These data points represented the physical position of the contour in space must be converted to pixel coordinates. For head-first supine positioning, the pixel coordinate (0, 0) is the upper left corner of the image. The (0, 0) in physical space is located at the center of the image however. In order to convert the contour's XYZ data into pixel coordinates, a shift is applied in order

to translate the physical (0, 0) to the upper left corner. The coordinates are then divided by the inter-pixel distance value set in the DICOM image header.

After the segment conversion is complete, the image information inside the GTV must be extracted. This can be performed using the function *poly2mask* in MATLAB. This function automatically creates a binary mask from a polygon defined by a set of points on a plane. Pixels within the polygon are set to a value of 1, and pixels outside the polygon are set to 0. In order to determine if an edge pixel lies inside or outside the polygon, *poly2mask* generates 5-by-5 grid of pixels inside the edge pixel. If the central pixel in subpixel grid falls inside the border, the edge pixel is counted as inside. Several contours generated in FIRE were compared with contours generated in CERR, a common research environment also utilizing *poly2mask*, and were found to be the same.

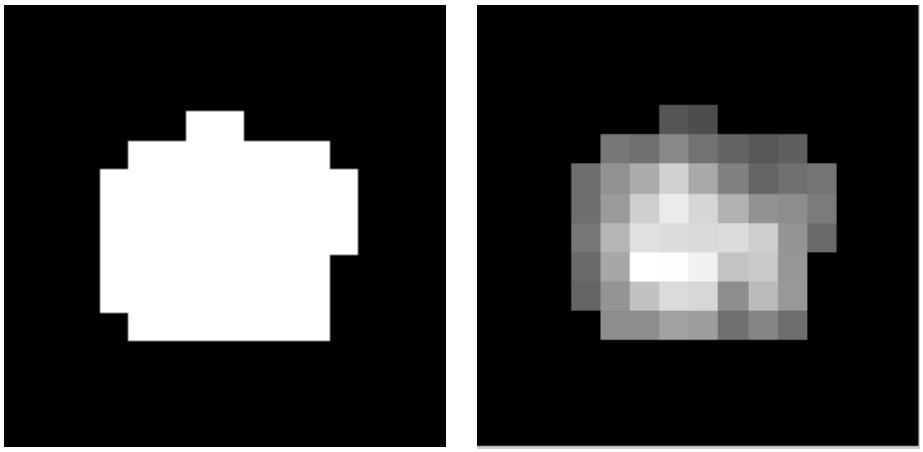


Figure 14: (left) The mask is generated using the *poly2mask* MATLAB function. This mask is then used to find the corresponding pixels from the image slice and (right) generate the extracted GTV.

After the masks are generated, any independent contour sections from a common volume occurring on the same slice are combined. The masks are then ordered by the position along the z-axis. The code then finds the original PET images with the corresponding contour mask. For any pixel that is of value 1 in the mask image, the corresponding pixel in the original image is saved to a new image consisting of an array of zeros as depicted in Figure 14. Thus, the code uses the mask to copy only the pixels inside the contour. With the newly created image of only the GTV, the SUV metrics and texture features can be calculated.

5.3 SUV, MV, and TLG Calculation

Duke FIRE is designed to be an inclusive program and therefore calculates standard uptake values, metabolic volume, and total lesion glycolysis. Several factors must be considered when performing this calculating, most importantly the decayed activity of the injected FDG. The only decay term accounted for in the raw DICOM data is the decay during acquisition. PET scan can take several minutes and as the scan time increases, the source will decay. Areas of the body scanned at the beginning versus the end may appear artificially high in activity when compared with the rest of the image. This time decay is automatically accounted for by the scanner. What is not accounted for is the decay in activity of the injected dose from the time of injection to the time of the scan. FDG takes up to an hour to disperse throughout the body, during which the

activity decreases. A decay term must be applied when calculating SUV from the raw image data. SUV is calculated according to Equation 10.

$$I_{SUV} = \frac{I \times RescaleSlope\ (\frac{Bq}{ml}) \times Body\ Weight\ (kg) \times 1000\ \left(\frac{g}{kg}\right)}{Radioactivity\ (Bq) \times e^{-t \times ln(2)/T_{1/2}}} \quad (10)$$

The calculation uses the previously mentioned RescaleSlope in order to establish units of Bq/ml. The exponential term on the denominator accounts for the radioactivity decay between the injection and scanned time, both of which are included in the DICOM header file. The half-life, $T_{1/2}$, of ^{18}F-FDG is 110 minutes.

Metabolic volume is calculated by multiplying the total number of voxels in the GTV by the pixel x-y widths and slice thickness found in the DICOM header file. Slice thickness is the space between image slices and represents the length of the voxel along the z-axis. Total lesion glycolysis is calculated by $MV \times SUV_{mean}$.

SUV calculated in FIRE matches closely SUV calculated from MIM. In a subset of 10 patients, SUV_{mean} was derived from both MIM and Duke FIRE. The median percent error of Duke FIRE compared to MIM was 1.53%. Small deviations may exist because MIM segments the edge pixels and only counts the volume inside the contour. Other errors may exist due to rounding in the process of converting the physical contour data to pixel contour data.

6. Texture Analysis of Gynecological Tumors

This study uses PET data from a prospective, IRB approved protocol enrolling a 29 women with node positive gynecological malignancies visible on PET: 15 with cervical cancer, 8 with uterine cancer, 1 with vaginal cancer, and 5 with vulvar cancer. A baseline PET-CT was acquired for treatment planning and normal tissues contoured. The PET primary GTV was contoured by an experienced radiation oncologist. Prescribed dose varied between 45-50.4 Gy, with a 55-70 Gy boost to the PET positive nodes. An intra-treatment PET-CT was acquired between 30 and 36 Gy. Tumor response was determined based on RECIST on a post-treatment PET-CT[43].

The purpose of this chapter is to determine the ability of texture analysis to assess tumor response in comparison to SUV histogram features, MV, and TLG. Chapter 6 covers 5 sections including the methods, results, and discussion of texture analysis of gynecological tumors via PET imaging. Section 6.1 discusses the statistical methods used to analyze the data. Section 6.2 examines the prognostic power of 2^{nd} and high order texture features compared with 1^{st} order features from baseline PET using PET Edge contoured primary tumors. Section 6.3 performs similar analysis from baseline PET using threshold contoured primary tumors. Section 6.4 explores the ability of 1^{st}, 2^{nd}, and high order features to track changes in tumor texture during treatment. Finally, section 6.5 correlates texture features with tumor physiology.

6.1 Statistical Analyses

A series of statistical algorithms were implemented through JMP statistical software (SAS Institute Inc., Cary, North Carolina) in order to assess the power of texture features from the baseline PET-CT primary GTV in predicting tumor response. The initial method of comparison and correlation of features with treatment outcome utilized the Wilcoxon rank-sum test.

The Wilcoxon rank-sum nonparametric test is used when comparing sets of unpaired data, such as data reflecting responders and non-responders as in this case, and produces a p-value. Representing the statistical significance of the separation of data, the p-value is a generally accepted classification metric in health sciences. Values lower than p = 0 .05 suggest a pattern exists between the data sets at or above the 95% confidence interval[44]. In this study, p-values indicate whether or not there is a reliable correlation between texture features and tumor response. These results are then compared with the rank-sum tests of SUV metrics including SUV_{max}, SUV_{mean}, skewness, and kurtosis as well as MV and TLG. Both the PET Edge and 40% threshold contours are tested in order to determine if the segmentation method significantly impacts texture features.

Receiver operating characteristic (ROC) curves were generated for both the PET Edge and threshold segmented data sets. An ROC curve is a method used to determine a threshold that optimizes the ability to distinguish between groups in a 2-class system.

The curve displays the balance between sensitivity (true positives) and false positives (1-specifictiy). High sensitivity indicates that the test can reliably predict the positive result while high specificity represents the ability of the test to distinguish negative results. The experimental positive is frequently the 'disease' condition, such as a positive diagnosis. In this study, the experimental positive classifier is no response (NR). Area under the curve (AUC) is a single metric used to assess the quality of the ROC curve. A perfect test would have an AUC = 1 with 100% sensitivity and 100% specificity. A failed test would have an AUC = 0.5 indicating that the test cannot reliably distinguish true or false positives[45].

Binomial logistic regression is a method frequently utilized in machine learning of classifying data into two categories, such as responder/ non-responder. Essentially, the result is a probability of data matching a classifier according to the independent variables. Probabilities are determined based on thresholds found using ROC analysis. Inverse prediction is utilized in determining threshold levels correlating to a designated probability. Logistic regression may be used in classifying data with no associated outcomes based on the probability[46].

6.2 Tumor Response Prediction using PET Edge Contoured Baseline GTVs

Of the 29 patients with primary tumor response evaluated via imaging, there were 16 CR, 7 PR, and 6 NR. Tumor response to therapy was determined by a physician using the post-treatment PET or CT and classified as either complete response (CR),

partial response (PR), or no response (NR) according to RECIST. Baseline primary PET positive GTVs were contoured using PET Edge and checked against the clinical GTVs contoured by a radiation oncologist. First, second, and high order features were extracted from the PET Edge contoured GTVs using Duke FIRE. Two groupings were assessed using Wilcoxon rank-sum tests correlating baseline features with tumor response: one comparing CR/PR vs. NR and one comparing CR vs. PR/NR. ROC analyses and logistic regression was performed for the CR/PR vs. NR group in order to assess the ability to classify responding and non-responding tumors using texture features.

6.2.1 Results from using PET Edge Contoured Baseline GTVs

From the CR/PR vs. NR group, 7 of the 40 texture features had a p-value < 0.05 in addition to skewness, kurtosis, and TLG. No features from the GLRLM were found to be statistically significant. From the CR vs. PR/NR group, 3 texture features including *energy* (p = 0.0364), *entropy* (p = 0.0240), and *max probability* (p = 0.0202) had a p-value < 0.05 in addition to kurtosis (p = 0.0169).

The CR/PR vs. NR group shows the highest potential for classifying PET GTVs as responders or non-responders. ROC analysis yielded an AUC > 0.8 for 4 texture features as well as kurtosis. All texture features presented an AUC > 0.75 while both SUV_{max} and SUVmean presented an AUC < 0.7. Logistic regression and inverse prediction tests were

performed for all metrics with $p < 0.05$ in order to determine predicted values for probabilities of 95%, 90% and 50%.

Table 3: Statistics for Texture Features with $p < 0.05$ and SUV Histogram Features of the PET Edge contoured Group CR/PR vs. NR.

Feature	p	AUC	Threshold	Sens. (%)	Spec. (%)	Prob. (%)
Energy	0.0077	0.862	.00165	83.3	69.6	24.17
Entropy	0.0066	0.870	6.63	83.3	73.9	21.54
Max Probability	0.0040	0.891	.00552	83.3	87.0	25.69
ZGLNU	0.0382	0.783	17.1	83.3	65.2	16.63
ZSNU	0.0335	0.790	313.6	83.3	65.2	14.85
Contrast (NGLDM)	0.0435	0.775	1.96×10^{-13}	83.3	69.6	16.90
Complexity	0.0166	0.826	1.47	83.3	69.6	20.80
Skewness	0.0382	0.783	.237	66.7	87.0	35.37
Kurtosis	0.0090	0.855	1.95	83.3	82.61	35.36
SUV_{max}	0.3461	0.623	19.0	83.3	45.5	18.12
SUV_{mean}	0.2697	0.645	8.94	83.3	56.5	18.78
MV	0.0713	0.746	55.8	83.3	65.2	17.42
TLG	0.0435	0.775	667.0	83.3	65.2	18.60

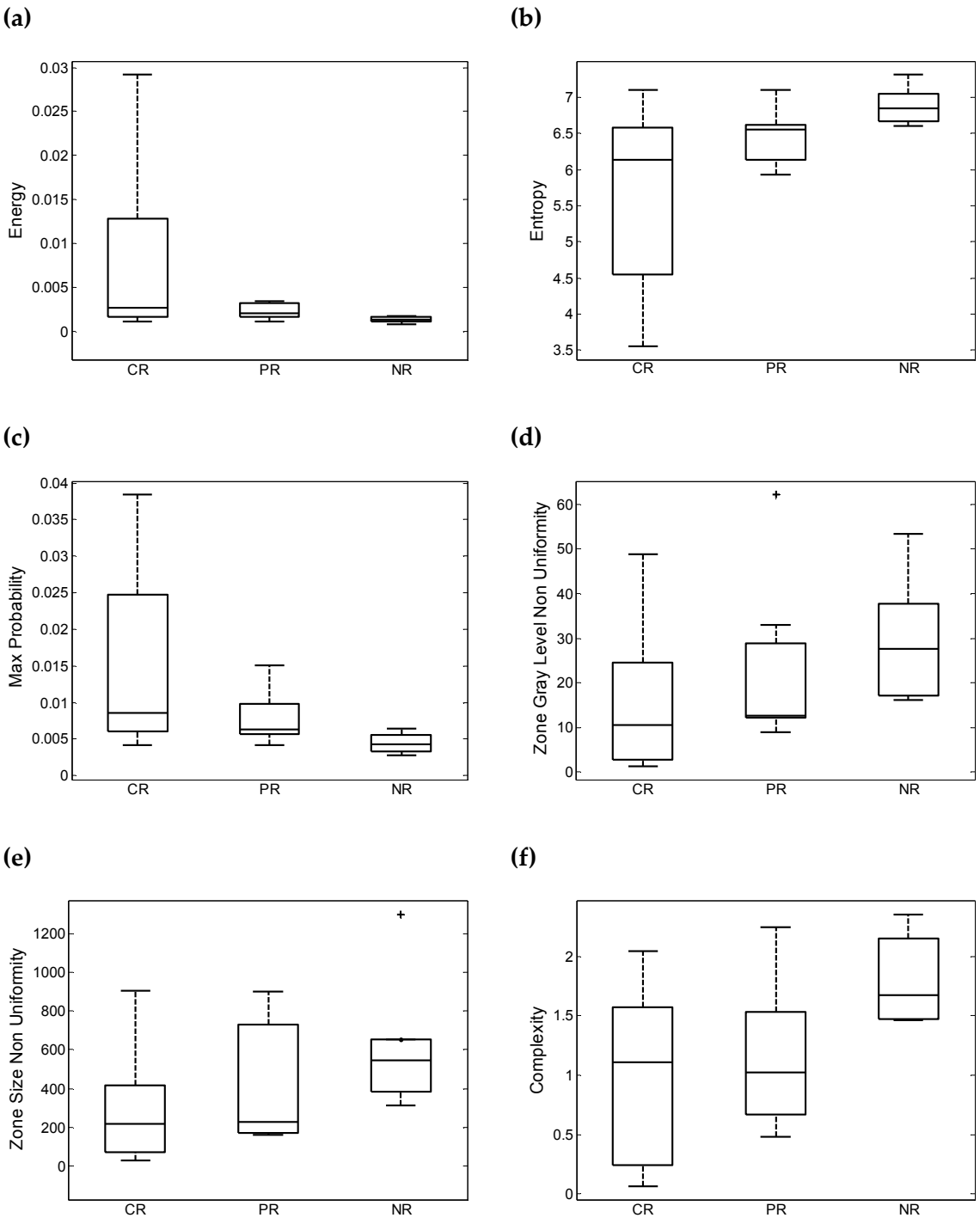


Figure 15: Boxplots comparing statistically significant 2nd and high order texture features derived from PET Edge contoured CR, PR, and NR tumors.

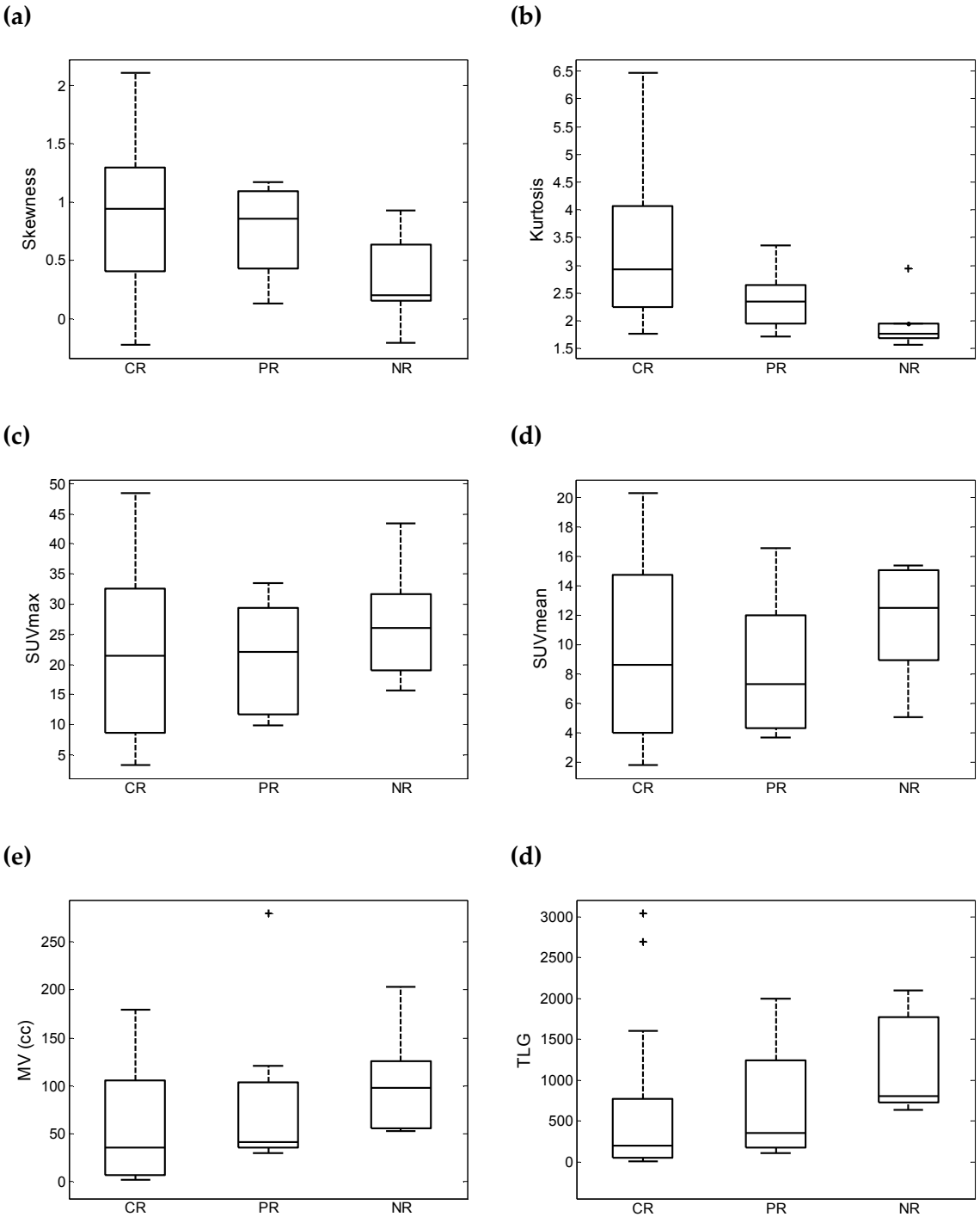


Figure 16: Boxplots comparing 1st order features derived from PET Edge contoured CR, PR, and NR tumors.

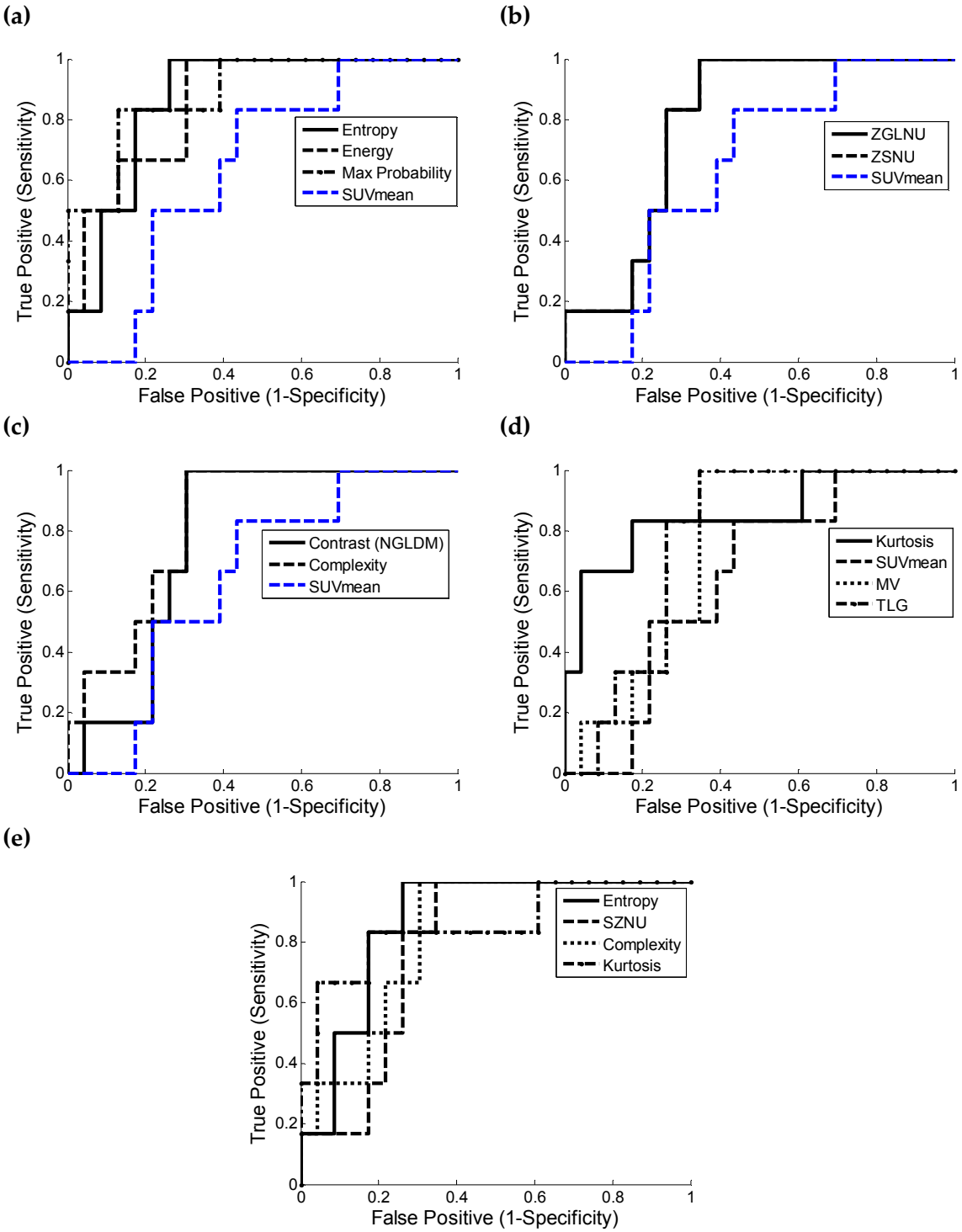
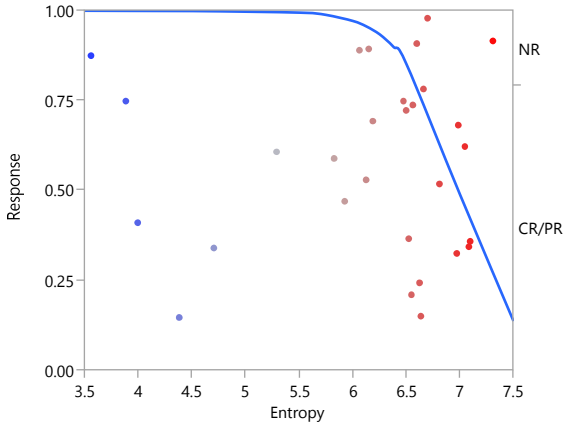


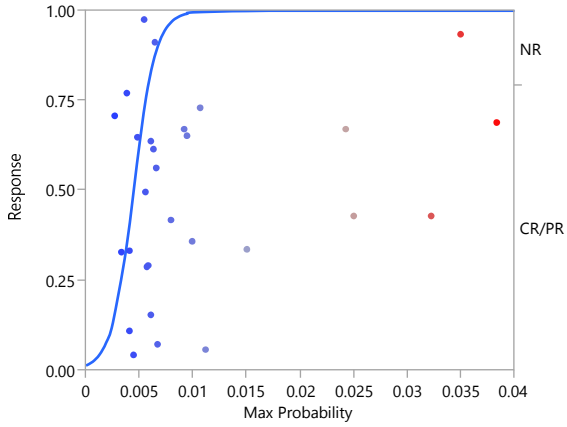
Figure 17: ROC curves comparing texture features from (a) the GLCM, (b) the GLSZM, and (c) the NGLDM against SUV_{mean}. (d) ROC curves of only 1st order features. (e) ROC curves comparing selected 1st, 2nd, and high order features.

(a)



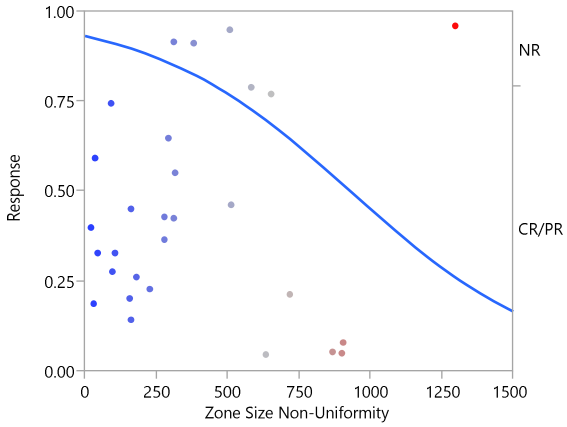
CR/PR Probability	Predicted Entropy	Lower 95%	Upper 95%
0.95	6.16	.	6.56
0.90	6.37	.	6.70
0.50	6.98	6.65	.

(b)



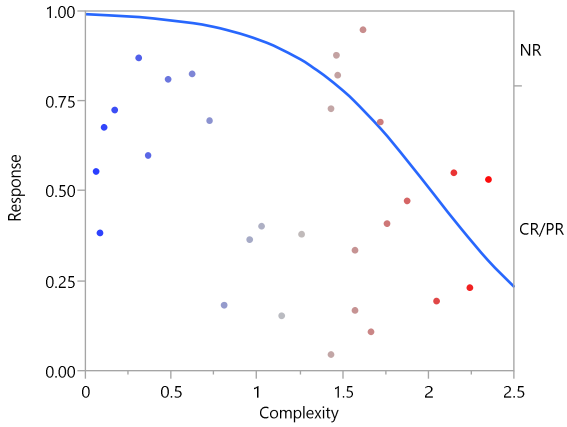
CR/PR Probability	Predicted Max Prob.	Lower 95%	Upper 95%
0.95	0.00746	0.00593	.
0.90	0.00669	0.00541	.
0.50	0.00443	.	0.00597

(c)



CR/PR Probability	Predicted ZSNU	Lower 95%	Upper 95%
0.95	-109	.	312
0.90	154	.	516
0.50	930	564	.

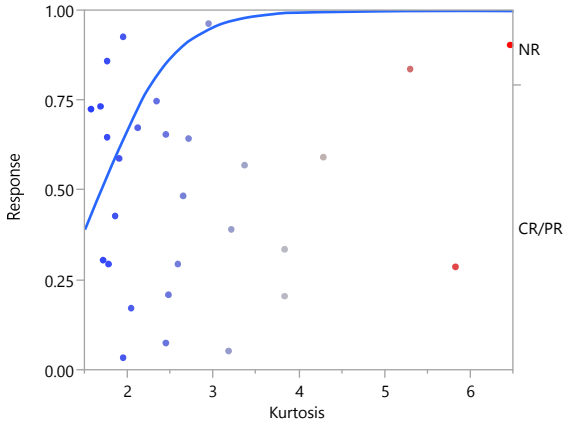
(d)



CR/PR Probability	Predicted Complexity	Lower 95%	Upper 95%
0.95	0.816	-475	1.35
0.90	1.12	-303	1.57
0.50	2.02	1.57	205

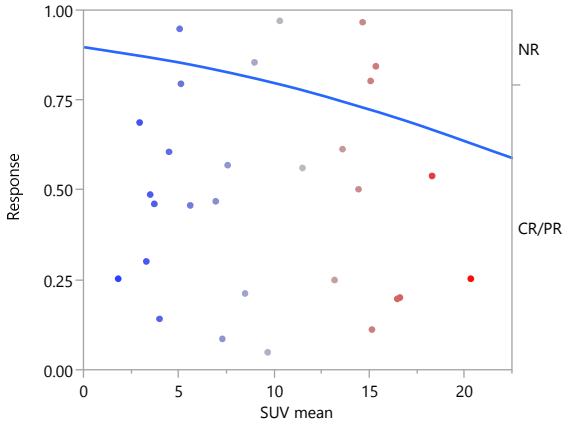
Figure 18: Logistic regression plots for 4 texture features with strong predictive power: (a) *entropy*, (b) *max probability*, (c) *ZSNU*, and (d) *complexity*. Each feature presents a model strong enough to inversely predict values for the 95%, 90%, and 50% probabilities of classifying the tumor as a CR/PR except *ZSNU*.

(a)



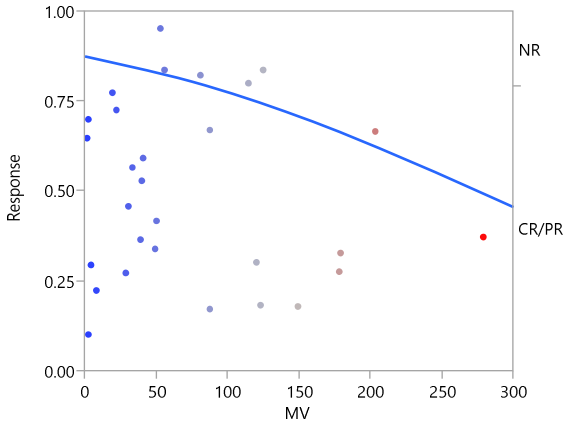
CR/PR Probability	Predicted Kurtosis	Lower 95%	Upper 95%
0.95	2.95	2.35	.
0.90	2.63	2.15	.
0.50	1.69	.	2.18

(b)



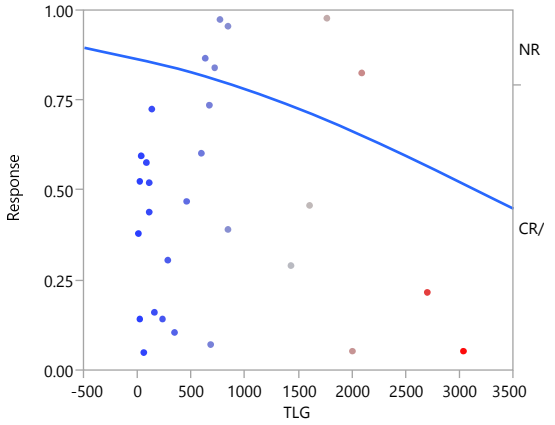
CR/PR Probability	Predicted SUV_{mean}	Lower 95%	Upper 95%
0.95	-9.29	.	6.41
0.90	-0.0846	.	11.8
0.50	27.0	13.8	.

(c)



CR/PR Probability	Predicted MV	Lower 95%	Upper 95%
0.95	-140	.	34.2
0.90	-34.8	.	102
0.50	276	127	.

(d)



CR/PR Probability	Predicted TLG	Lower 95%	Upper 95%
0.95	-1809	.	265
0.90	-549	.	1088
0.50	3158	1387	.

Figure 19: Logistic regression plots for kurtosis, SUV_{mean}, MV, and TLG. Kurtosis is the only model capable of predicting the 95%, 90%, and 50% probabilities of classifying the tumor as a CR/PR.

6.2.2 Discussion of Tumor Prediction using PET Edge Contoured Baseline GTVs

Three features from the GLCM appear as statistically significant classifiers of responders and non-responders: *max probability, energy*, and *entropy*. *Max probability* presents the strongest prediction power out of all texture features and is defined as the highest value in the co-occurrence matrix. Uniform images tend to have more of the same pairs of pixels resulting in larger values, but fewer combinations, in the co-occurrence matrix, P(i,j). Note that it is not necessary for pixels in the pair to have the same intensity to be considered homogeneous. Rather, only a large number of the same pairs of pixels must appear.

$$Max\ Probability = Max\ P(i,j) \tag{11}$$

Figure 15.c shows a clear grouping between CR, PR, and NR with larger values corresponding to more positive prognosis. This pattern categorizes *max probability* as a measurement of homogeneity similar to *energy*.

$$Energy = \sum_i \sum_j P(i,j)^2 \tag{12}$$

Energy is weighted towards homogenous images by summing the square of the co-occurrence matrix contents. Both *energy* and *max probability* are closely proportional to the maximum value in the matrix while *entropy* is inversely related to the maximum matrix value due to the logarithm.

$$Entropy = \sum_i \sum_j P(i,j) \log P(i,j) \tag{13}$$

High *entropy* signifies many pairs of different pixel intensities within the co-occurrence matrix. In order for *entropy* to be large, P(i,j) must remain relatively small. *Max probability, energy*, and *entropy* are closely correlated to the maximum value within the co-occurrence matrix and should therefore be as a whole rather than three independent features as they all likely reflect the same texture characteristics.

ZGLNU and *ZSNU* derived from the GLSZM appear as significant classifiers, though weaker than the GLCM features. *ZGLNU* measures the similarity in pixel intensities throughout the image and *ZSNU* measures the similarity in zone sizes as both sum the squares of the zone size matrix contents, weighting the measurement to the largest homogenous zone. The similarity in equations indicates that the two features are closely related.

$$ZGLNU = \frac{1}{N_r}\sum_{i=1}^{M}\{\sum_{j=1}^{N} P(i,j)\}^2 \tag{14}$$

$$ZSNU = \frac{1}{N_r}\sum_{j=1}^{N}\{\sum_{i=1}^{M} P(i,j)\}^2 \tag{15}$$

Figure 15.d and 15.e show *ZGLNU* and *ZSNU* as increasing in PR and NR tumors. This correlation may not appear to be an intuitive outcome initially as *energy* was also a sum of squares, yet correlated with homogeneity. The difference is in identifying homogenous textures as distributions in which the spatial rate of change in pixel intensities is consistent. In other words, the distribution of pixel intensities

includes many small, repeating patterns. Heterogeneous images tend to have larger zones of very different intensity pixels. Equation 14 and Equation 15 show that larger zones lead to increased heterogeneity and a poorer prognosis. *ZSNU* shows slightly lower p-value and higher AUC than *ZGLNU*, however, both features are closely correlated. These features likely describe the same textures and should not be considered as independent from each other.

The GLSZM is based off the same principles as the GLRLM, yet no features appear statistically significant from the GLRLM. A possible reason is that run lengths features do not show strong prognostic power is that lengths of pixels may not correlate with physiology of aggressive or more benign tumors. On the other hand, the GLSZM tabulates 3D regions of uniform pixels within the tumor, possibly identifying hypoxic or necrotic areas that correlate with poor prognosis.

Contrast and *complexity* are more difficult to connect with underlying physiology. Both reflect the information content of an image and will increase with more varied pixel intensity information. *Contrast* defined in the context of the NGLDM represents the apparent difference between neighboring regions of pixel intensities. It is the squared difference between pixel pairs weighted by the probability of the pairs occurring and scaled by the average neighborhood intensity value.

$$Contrast = \left[\frac{1}{N_g(N_g-1)}\sum_{i=0}^{G_h}\sum_{j=0}^{G_h} p_i\, p_j\, (i-j)^2\right]\left[\frac{1}{n^3}\sum_{i=0}^{G_h} s(i)\right] \tag{16}$$

Complexity describes the information of pixel distributions in an image and shows potential for delineating responders from non-responders. The neighborhood difference matrix samples local 3D 'patches' of the image. If there are many patches of different intensities, complexity will be high. This is the case when images have a high rate of spatial intensity change. More uniform images in which the patches show less variation will have a lower complexity.

$$Complexity = \left\{\sum_{i=0}^{G_h}\sum_{j=0}^{G_h}\frac{|i-j|}{n^3(p_i+p_j)}\right\}\{p_i s(i) + p_j s(j)\} \tag{17}$$

$$p_i \neq 0, p_j \neq 0$$

Non-responding tumors tend to have more texture, thus, *contrast* and *complexity* are measures of heterogeneity increasing with poorer prognosis. It is interesting to note that nearly all the statistically significant texture features tend to indicate a pattern, such as increasing or decreasing, between CR, PR, and NR. *Complexity* presents a slightly different case. Figure 15.f shows a very slight decrease in median complexity from CR to PR, then an increase in NR. Understanding the underlying cause behind this is difficult and only inference can be made. *Complexity* may be less sensitive to small changes in tumor structure, such as the small increase in heterogeneity from CR to PR, and may require a threshold level of texture information to differentiate between groups.

SUV histogram features are significantly weaker predictors of tumor response in this study than texture features. SUV_{max} and SUV_{mean} show no power in differentiating responders from non-responders, although skewness and kurtosis are statistically significant metrics. TLG is statistically significant despite both SUVmean and MV low p-values and should be considered as a metric in future studies.

The AUC, threshold, sensitivity, specificity, and probability values given in Table 3 are additional statistical descriptors of the data. The probabilities for each feature represent the probabilities of classifying the tumor as a non-responder at the specific threshold with the associated sensitivity and specificity. The 2nd order features, *energy*, *entropy*, and *max probability*, have the highest AUC with the best balance between sensitivity and specificity of all texture features. Figure 17 indicates that texture features are generally better at classifying responders and non-responders than SUV metrics.

Figure 18 presents promising results from logistic regression analysis. In particular, Figure 18.a and Figure 18.b show a strong separation between CR/PR and NR. The 1st order features in Figure 19 have much weaker separation however. In addition to the logistic regression plots, predicted values at given probabilities derived from inverse prediction are presented. Inverse prediction uses logistic regression to determine thresholds of features correlating with probabilities of a given outcome. In this case, the positive response was CR/PR and probabilities were set at 95%, 90%, and 50%. Determining values of features which can lead to a given probability of predicting

a CR/PR is useful in assessing incoming patient data with no associated outcome. This strategy is used in machine learning for testing if a training model holds true. Features tested with inverse prediction should yield a positive value correlating with each percentage if the model is strong. For example, inverse prediction of *entropy* ($p = 0.0077$) comparing CR/PR vs. NR yields a value of 6.17 correlating with 95%. Therefore, if a baseline PET GTV has an entropy value of 6.17 or lower, there is a 95% chance that the tumor will respond to radiation therapy. *Entropy* provides a strong model for predicting response, however, inverse prediction may also indicate if a model is weak.

The inverse prediction algorithm must extrapolate the data to find a threshold value correlating with a probability if there is decreased separation between the classifiers such as the SUV metrics. SUV_{mean} ($p = 0.270$) for example is a very weak classifier of tumor response. Inverse prediction therefore results in negative values for the 95% and 90% probabilities as is the case for MV ($p = 0.0713$), TLG ($p = 0.435$), and the 95% level of *ZSNU* ($p = 0.0355$). Negative values indicate that the model is not robust enough to reliably classifying data into responders or non-responders even if the nonparametric tests show statistical significance.

6.3 Tumor Response Prediction using 40% SUV_{max} Threshold Contoured Baseline GTVs

Baseline primary GTVs were contoured using via 40% threshold and checked against the clinical GTVs contoured by a radiation oncologist. First, second, and high order features were extracted from the baseline PET threshold segmented GTVs using

Duke FIRE and grouped into CR/PR vs. NR in order to classify between responding and non-responding tumors. Statistics including nonparametric testing, ROC, and logistic regression were performed in order to assess the ability of thresholding to yield reliable prediction models. Results from PET Edge and threshold segmentation methods are then compared and the differences between the tests are discussed.

6.3.1 Results from using 40% SUV_{max} Threshold Contoured Baseline GTVs

Only 5 texture features presented $p < 0.05$ compared to the 7 texture features with significantly low p-values derived from PET Edge segmentation. All p-values increased with threshold segmentation as compared to PET Edge segmentation. Additionally, AUC values decreased for all features except SUV_{max} which increased from 0.623 to 0.652 and TLG which increased from 0.775 to 0.783.

Table 4: Statistics for Texture Features with $p < 0.05$ and SUV Histogram Features of the 40% SUV_{max} threshold contoured Group CR/PR vs. NR.

Feature	P	AUC	Threshold	Sens. (%)	Spec. (%)	Prob. (%)
Energy	0.0255	0.804	.00201	83.3	73.9	29.86
Entropy	0.0382	0.783	6.42	83.3	73.9	26.74
Max Probability	0.0255	0.804	.00669	83.3	73.9	27.63
ZSNU	0.0382	0.692	243.0	83.3	60.9	20.19
Complexity	0.0494	0.768	.921	76.9	80.0	20.64
SUV_{max}	0.270	0.652	23.2	66.7	65.2	20.10
SUV_{mean}	0.319	0.638	12.2	66.7	53.3	19.56
MV	0.125	0.710	43.9	66.7	65.2	19.56
TLG	0.0382	0.783	400.3	83.3	69.6	17.59

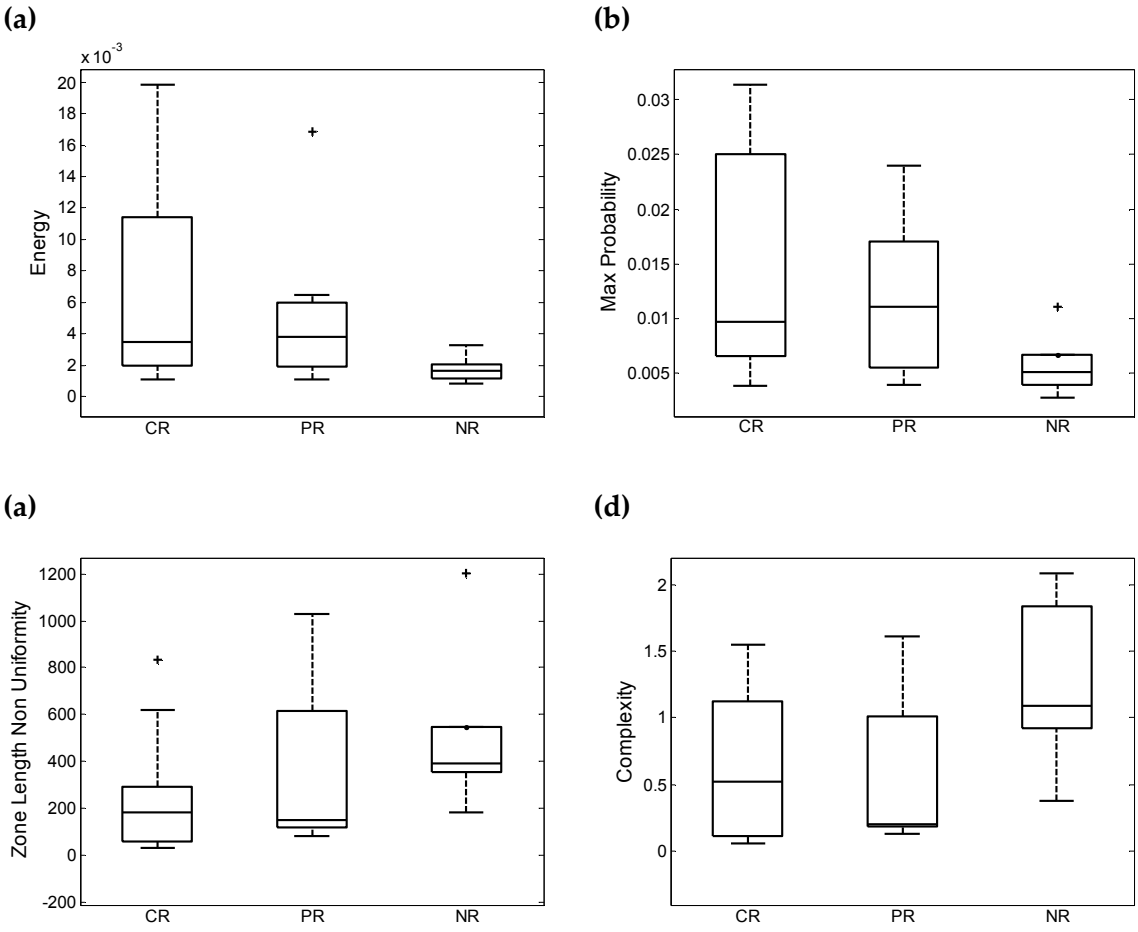


Figure 20: Boxplots comparing statistically significant 2nd and high order texture features derived from 40% threshold contoured CR, PR, and NR tumors.

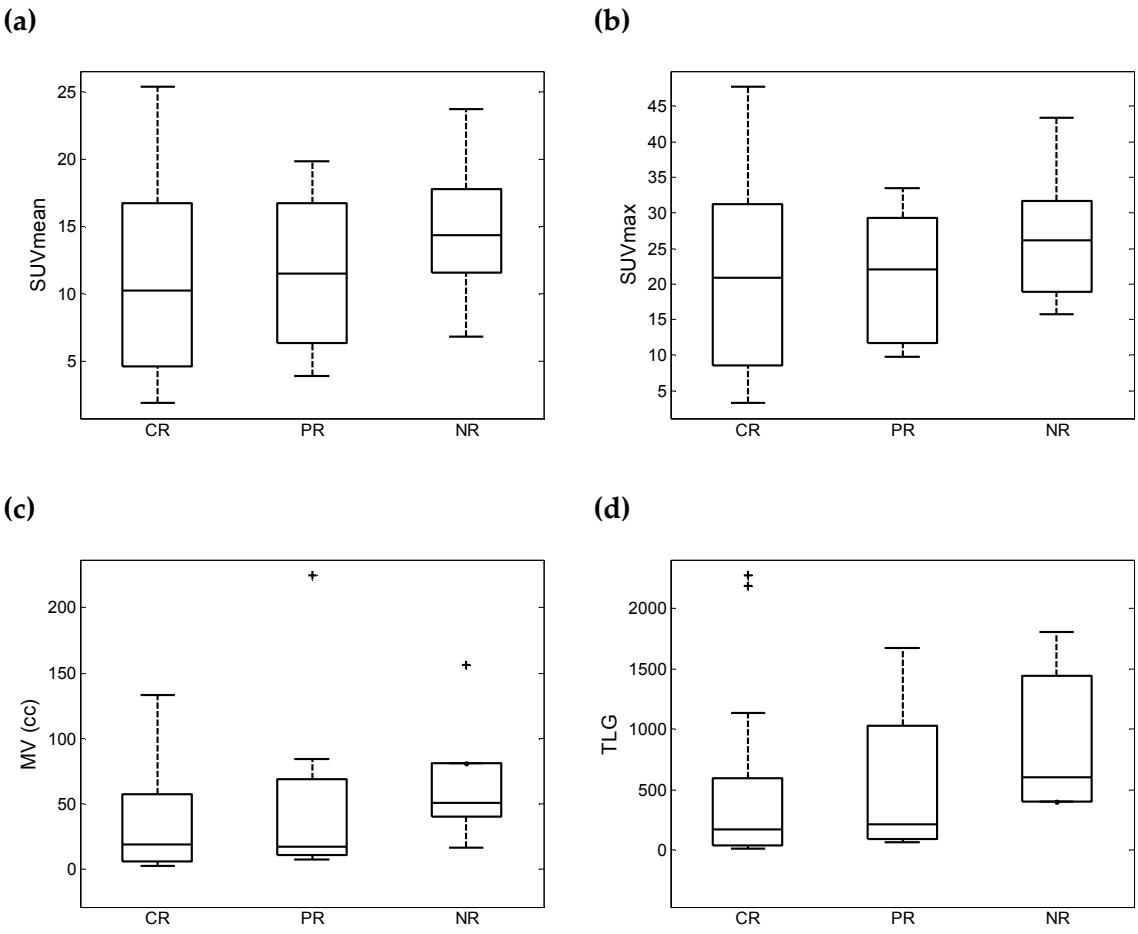


Figure 21: Boxplots comparing 1st order features derived from the 40% threshold contoured CR, PR, and NR tumors.

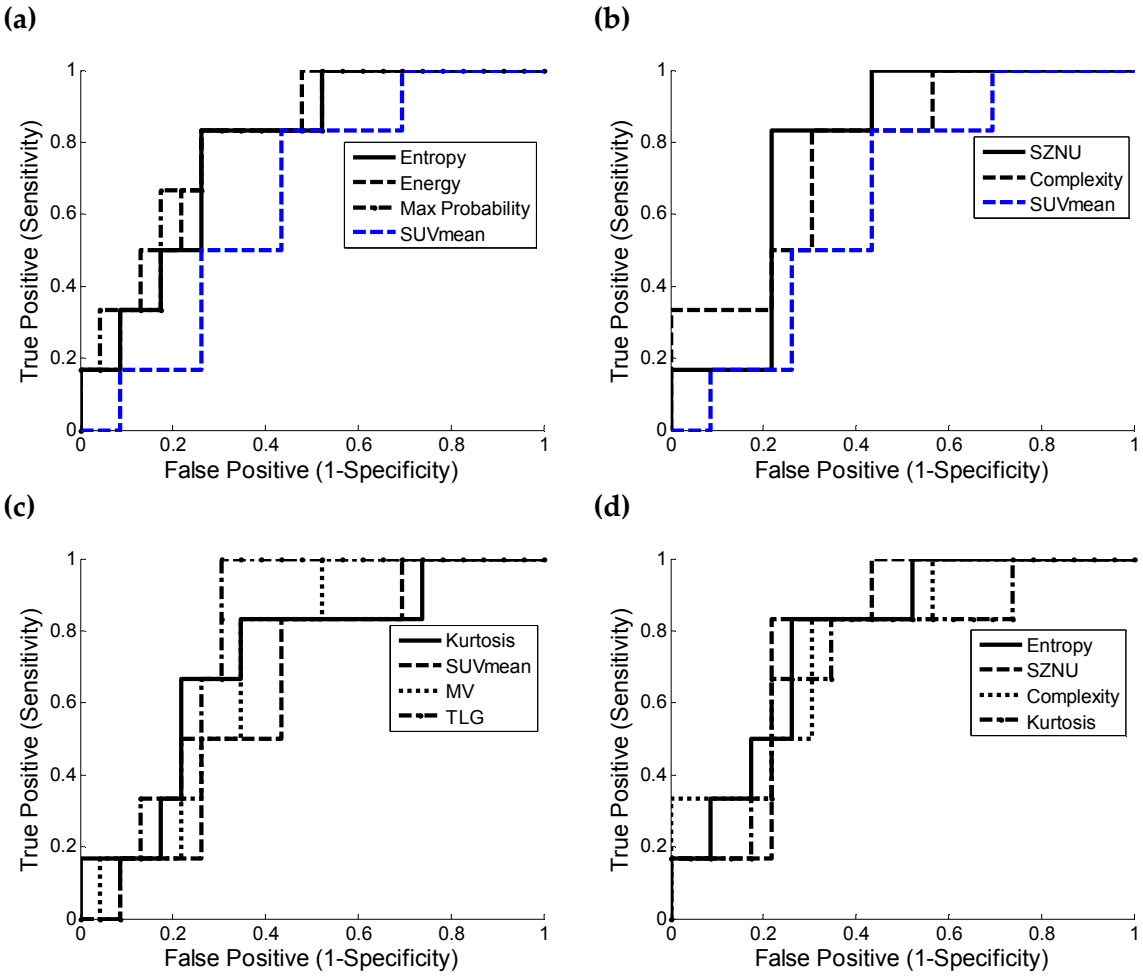
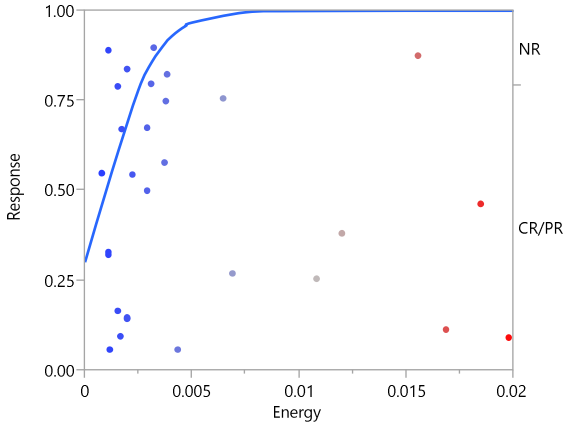


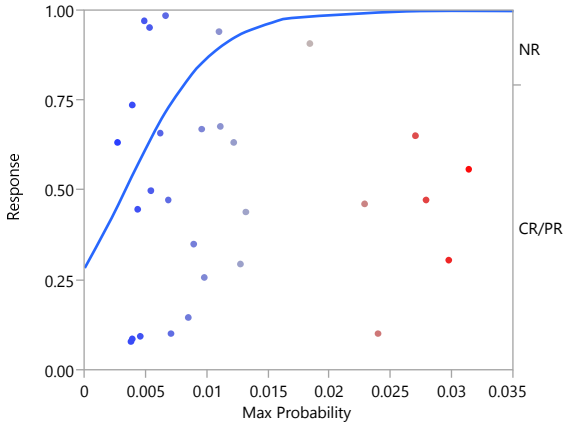
Figure 22: ROC curves comparing texture features from (a) the GLCM and (b) the GLSZM and NGLDM against SUV_{mean}. (c) ROC curves of only 1st order features. (d) ROC curves comparing selected 1st, 2nd, and high order features.

(a)



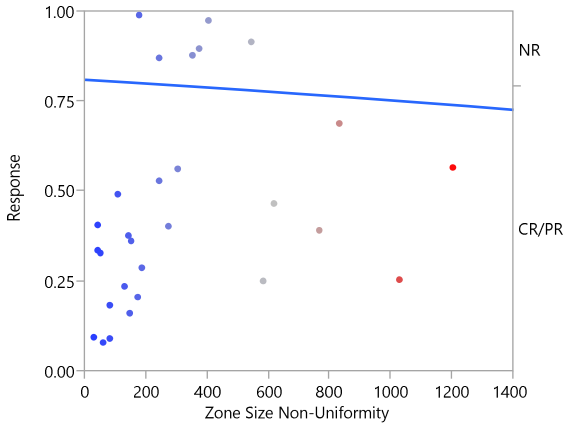
CR/PR Probability	Predicted Energy	Lower 95%	Upper 95%
0.95	0.0045	0.00278	.
0.90	0.0036	0.00227	.
0.50	0.0010	.	0.00241

(b)



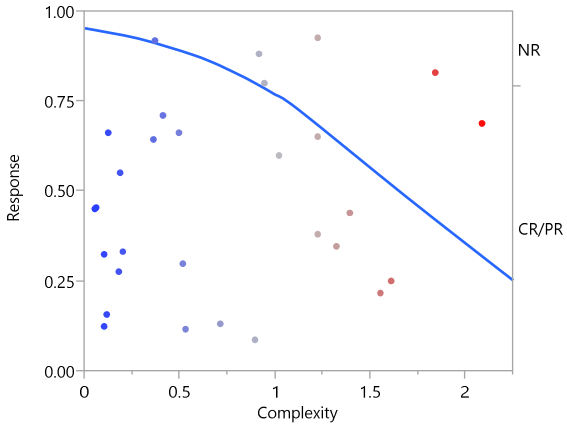
CR/PR Probability	Predicted Max Prob.	Lower 95%	Upper 95%
0.95	0.0137	0.00880	.
0.90	0.0111	0.00720	.
0.50	0.00328	.	0.00713

(c)



CR/PR Probability	Predicted ZSNU	Lower 95%	Upper 95%
0.95	-4385	.	-38.9
0.90	-2182	-24481	20116
0.50	4295	618	.

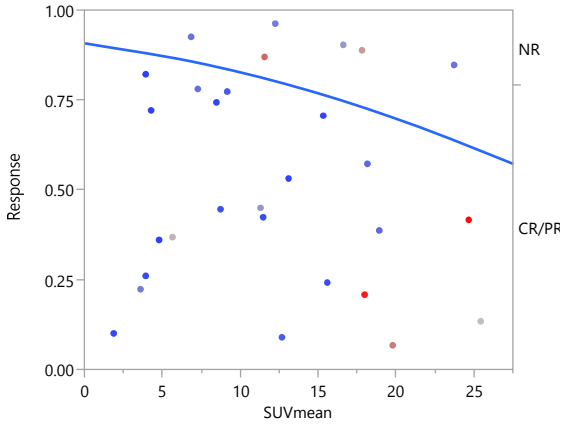
(d)



CR/PR Probability	Predicted Complexity	Lower 95%	Upper 95%
0.95	0.0491	.	0.718
0.90	0.457	.	1.03
0.50	1.66	1.09	.

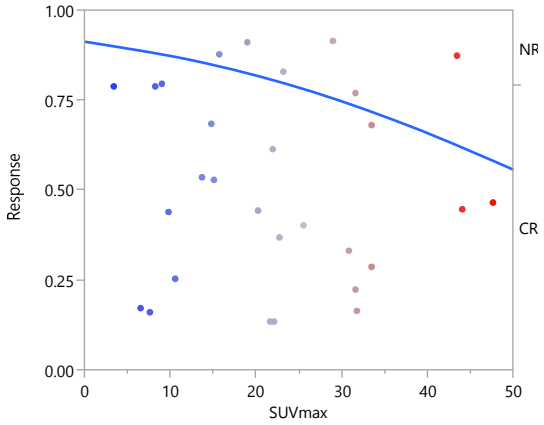
Figure 23: Logistic regression plots for 4 texture features with predictive power: (a) *energy*, (b) *max probability*, (c) *ZSNU*, and (d) *complexity*. These features provide weaker models for inversely predicting values for the 95%, 90%, and 50% probabilities of classifying the tumor as a CR/PR when compared to the PET Edge derived models. *ZSNU* fails both the 95% and 90% probabilities.

(a)



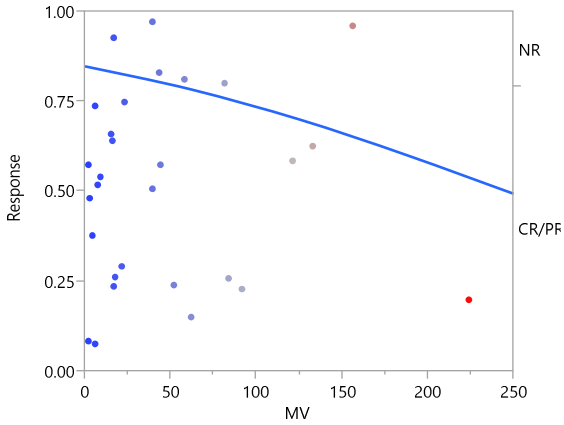
CR/PR Probability	Predicted SUV_{mean}	Lower 95%	Upper 95%
0.95	-8.75	.	8.44
0.90	1.48	.	14.7
0.50	31.6	17.1	.

(b)



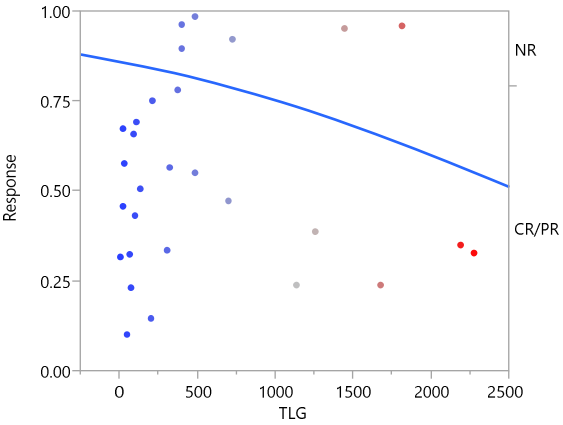
CR/PR Probability	Predicted SUV_{max}	Lower 95%	Upper 95%
0.95	-13.5	.	15.8
0.90	3.99	.	26.6
0.50	55.5	30.8	.

(c)



CR/PR Probability	Predicted MV	Lower 95%	Upper 95%
0.95	-176	.	6.73
0.90	-68.4	.	71.6
0.50	247	91.0	.

(d)



CR/PR Probability	Predicted TLG	Lower 95%	Upper 95%
0.95	-1569	.	166
0.90	-518	.	845
0.50	2573	1094	.

Figure 24: Logistic regression plots for SUV_{mean}, SUV_{max}, MV, and TLG. All 1st order features are weak models for probabilities of classifying the tumor as a CR/PR.

6.3.2 Discussion of Tumor Prediction using 40% SUV_{max} Threshold

Gradient-based segmentation results in more powerful texture features than thresholding for a very intuitive reason. Texture analysis depends on the information content within an image, but thresholding excludes portions of the information based on pixel intensity. A common physiological phenomenon correlating with poor prognosis are areas of low FDG uptake, as depicted in Figure 25, due to necrosis or hypoxia[47].

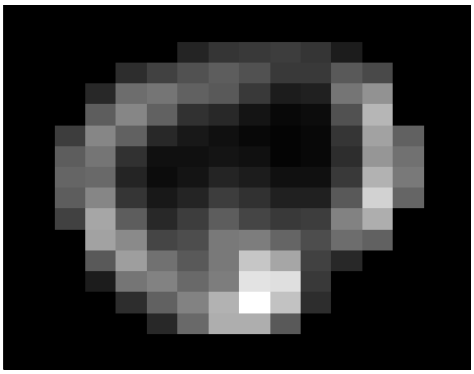


Figure 25: Primary tumor with a central area of low uptake. This tumor did not respond to radiation therapy.

Hypoxia in cervical tumors has been shown to be associated with increased malignance[48]. Yet thresholding will likely avoid including these areas in the segment. Using a gradient-based method such as PET Edge also for semi-automatic while maintaining the integrity of the entire GTV, including areas of low uptake and preserving the information content of the image.

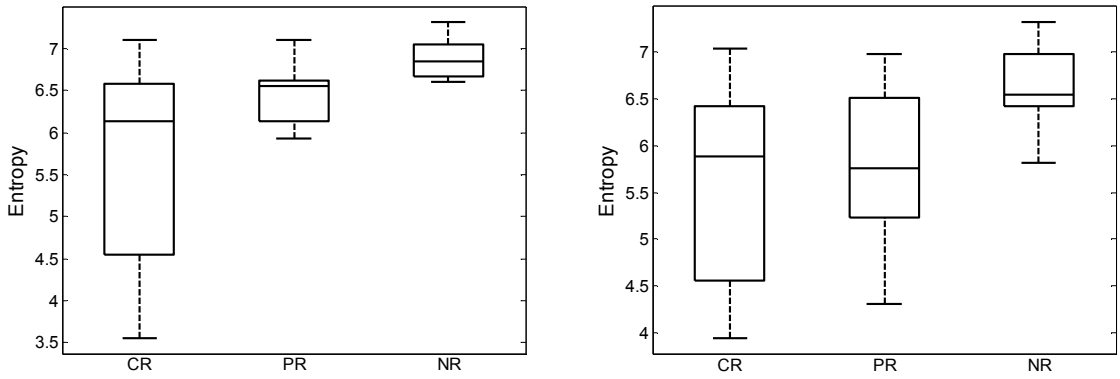


Figure 26: (left) PET Edge and (right) threshold boxplots of CR, PR, and NR.

Figure 26 presents a comparison of entropy from the PET Edge and threshold methods. The data regarding CR is nearly identical between the two data sets indicating that for tumors completely responding to radiation therapy, the segmentation technique has little impact. Intuitively, CR tumors will have a more homogeneous distribution of FDG uptake and therefore the threshold segmentation technique preserves most of the image content. A large difference occurs with PR tumors. As expected, partially responding tumors have slightly higher *entropy* values in the PET Edge data set because more texture is associated with poorer outcome. Thresholding may exclude the increase in texture however, if the texture is low uptake and the texture algorithm will understand the image information as more benign. Finally, the grouping of NR tumors in the PET Edge data set is confined to 6.5 – 7 while in the thresholding set, the values drop below 6.

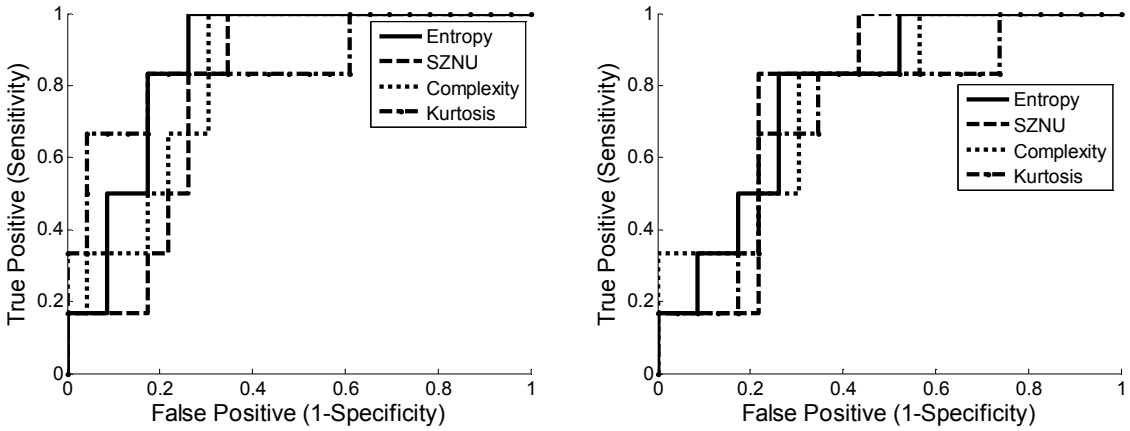
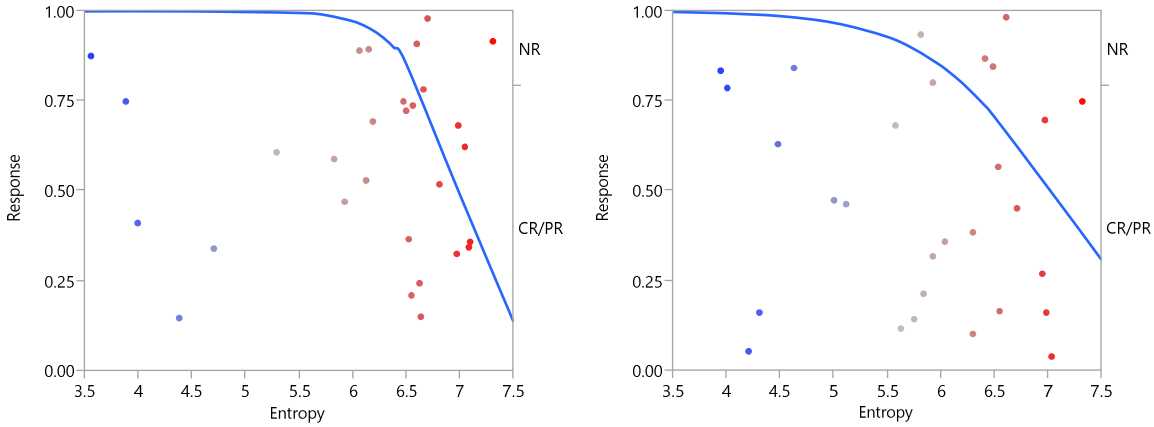


Figure 27: (left) PET Edge and (right) threshold ROC curves of selected 1st, 2nd, and high order features.

Figure 27 compares the ROC curves of PET Edge and threshold segmented GTVs. All features from the PET Edge GTVs show a higher AUC than the threshold method while the sensitivities and specificities between the features are comparable. Kurtosis was not statistically significant in the threshold model and is only shown for comparison.

Logistic regression and inverse prediction is performed on the threshold segmented features in order to determine the ability to predict tumor response. In general, the regression curves are less vertical indicating a decreased separation between responding and non-responding features.



CR/PR Probability	Predicted Entropy	Lower 95%	Upper 95%
0.95	6.17	.	6.56
0.90	6.38	.	6.70
0.50	6.99	6.65	.

CR/PR Probability	Predicted Entropy	Lower 95%	Upper 95%
0.95	5.27	.	6.05
0.90	5.72	.	6.35
0.50	7.02	6.39	.

Figure 28: (left) PET Edge and (right) threshold logistic regression plots.

Figure 28 compares the logistic regression plots and probabilities of CR/PR between *entropy* derived from PET Edge and threshold methods and shows that the PET Edge method provides a stronger model for classifying tumors. The predicted *entropy* values for the 95% and 90% probabilities were lower in the threshold model, however, the difference between the predicted *entropy* and upper 95% *entropy* were greater than the PET Edge model. Additionally, the PET Edge regression curve depicts a more favorable degree of separation between responders and non-responders than the threshold curve.

Factors such as segmentation can drastically affect results and though some papers have explored varying segmentation methods, many rely on thresholding to 40% of SUV_{max}. Segmentation plays a critical role because texture features are more sensitive to local and regional, rather than global, pixel distributions. If a region of pixels

representing tumor is not included in the ROI, the image loses valuable texture information whereas SUV_{mean} may only change minimally. By design, thresholding excludes areas of relatively low signal. Low signal represents low FDG uptake, potentially signifying an area of necrosis or hypoxia. Necrosis occurs when cells in a region of the tumor die and is commonly associated with poor outcomes. Hypoxia occurs when blood flow, and therefore oxygen, decreases in a region of tumor cells. If thresholding is used, these regions may be excluded and the texture matrices will not classify these image characteristics, losing some ability to detect non-responding tumors. The advantage of thresholding is reproducibility. Studies using thresholding are more easily comparable as it is a familiar segmentation method. If thresholding is shown to be inferior to a method such as semi-automatic gradient-base segmentation as used in this paper, or even manual segmentation, the reproducibility of thresholding is not worth the loss of texture information.

6.4 Temporal Changes in 1st, 2nd, and High Order Features

Having examined the prognostic ability of texture features derived from baseline PET scans, the next task is to determine how texture features reflect temporal changes in tumors such as over the course of treatment. Positive response to radiation therapy is often reflected by a decrease in SUV signifying fewer metabolically active cancer cells, or decreased metabolic activity of, roughly, the same cells. Decrease in tumor volume is also a positive sign of therapy, but neither SUV or nor MV indicate changes in tumor

heterogeneity which could reflect tumor outcome. Increased tumor heterogeneity on the baseline PET-CT is associated with poor prognosis, but a large shift towards increased homogeneity may be more indicative of tumor response than a change in SUV.

Of the 29 women with node positive gynecological malignancies, 28 had both baseline and intra-treatment PET-CT scans. The primary GTVs on both baseline and intra-treatment scans were contoured using PET Edge and checked against the clinical GTVs contoured by a radiation oncologist. Due to constraints of size in texture PET positive node GTVs were not considered for temporal changes. First, second, and high order features presenting statistically significant changes during therapy were determined by comparing CR, PR, and NR baseline and intra-treatment metrics using a Wilcoxon signed-rank test. The total patient group was than separated into responders (CR/PR) and non-responders (NR) in order to understand patterns that may reflect treatment outcome. Percent differences in all features from baseline to intra-treatment were found to identify features showing larger changes in responders than non-responders.

6.4.1 Results of Temporal Changes in 1st, 2nd, and High Order Features

A total of 14 texture features showed a statistically significant change ($p < 0.05$) from baseline to intra-treatment PET-CT. Six texture features presented a change with $p < 0.01$. SUV_{max}, SUV_{mean}, and TLG were the most significant indicators of tumor response

during treatment ($p < 0.00001$). Table 4 presents the signed-rank test results of the grouped CR, PR, and NR temporal data.

Table 5: 1st, 2nd, and high order features with $p < 0.05$. The features in bold are features that show promise as prognostic factors. *Max probability* and *Contrast (NGLDM)*, features that appeared to have prognostic significance, do not appear to correlate with tumor response during therapy.

Energy: p = 0.0427	*LZHGLE*: p = 0.0228	***Complexity***: p = 0.0015
Entropy: p = 0.0077	***ZGLNU***: p < 0.001	*Texture Strength*: p = 0.0288
RLGLNU: p = 0.0242	***ZSNU***: p = < 0.001	SUV_{max}: p < 0.00001
RLNU: p < 0.001	*Zone Percentage*: p = 0.0232	SUV_{mean}: p < 0.00001
Run Percentage: p = 0.0355	*Coarseness*: p = .0305	MV: p < 0.001
SZE: p = 0.0382	*NGLDM Contrast*: p < 0.001	TLG: p < 0.00001

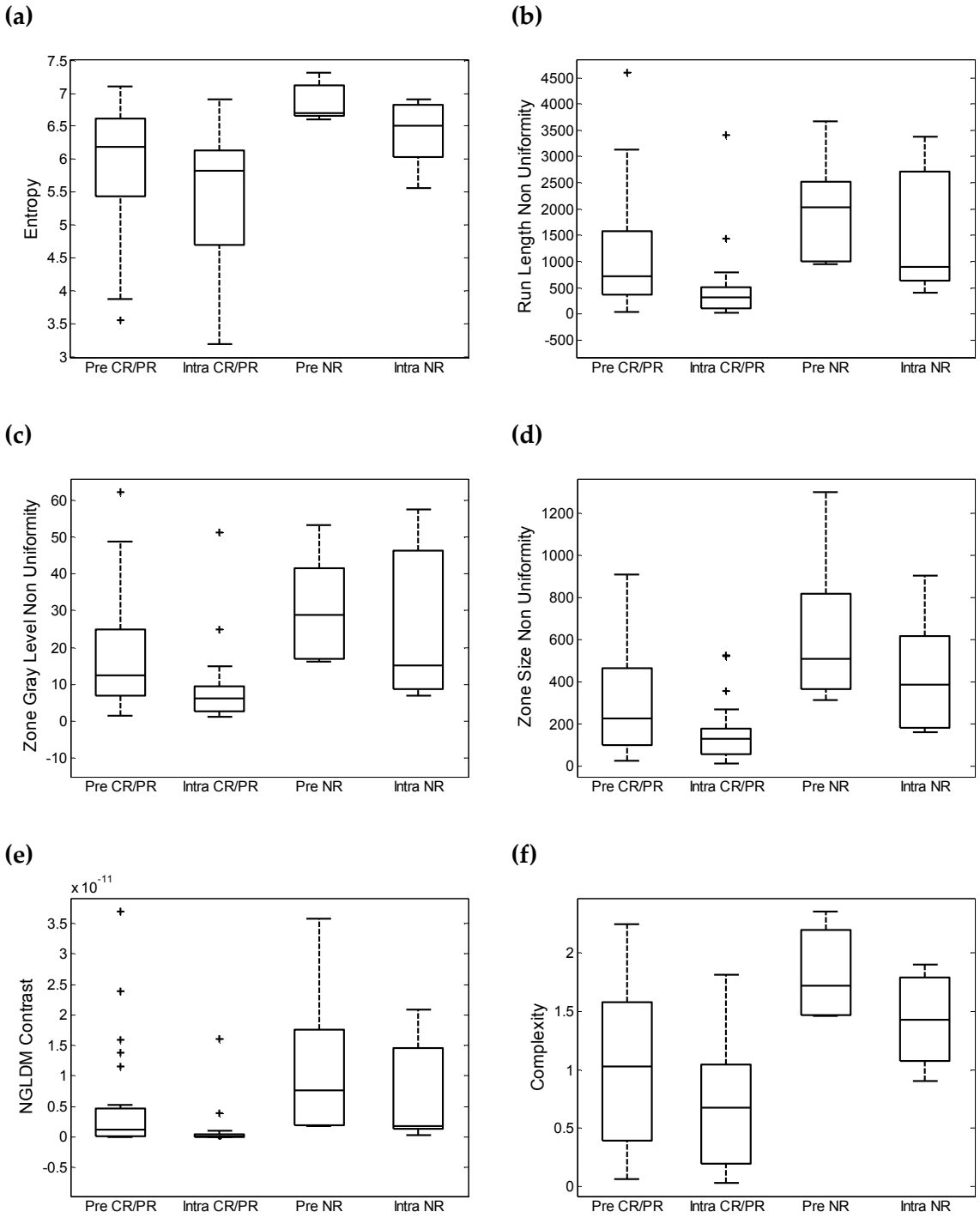


Figure 29: Comparison of baseline and intra-treatment texture features split between CR/PR and NR.

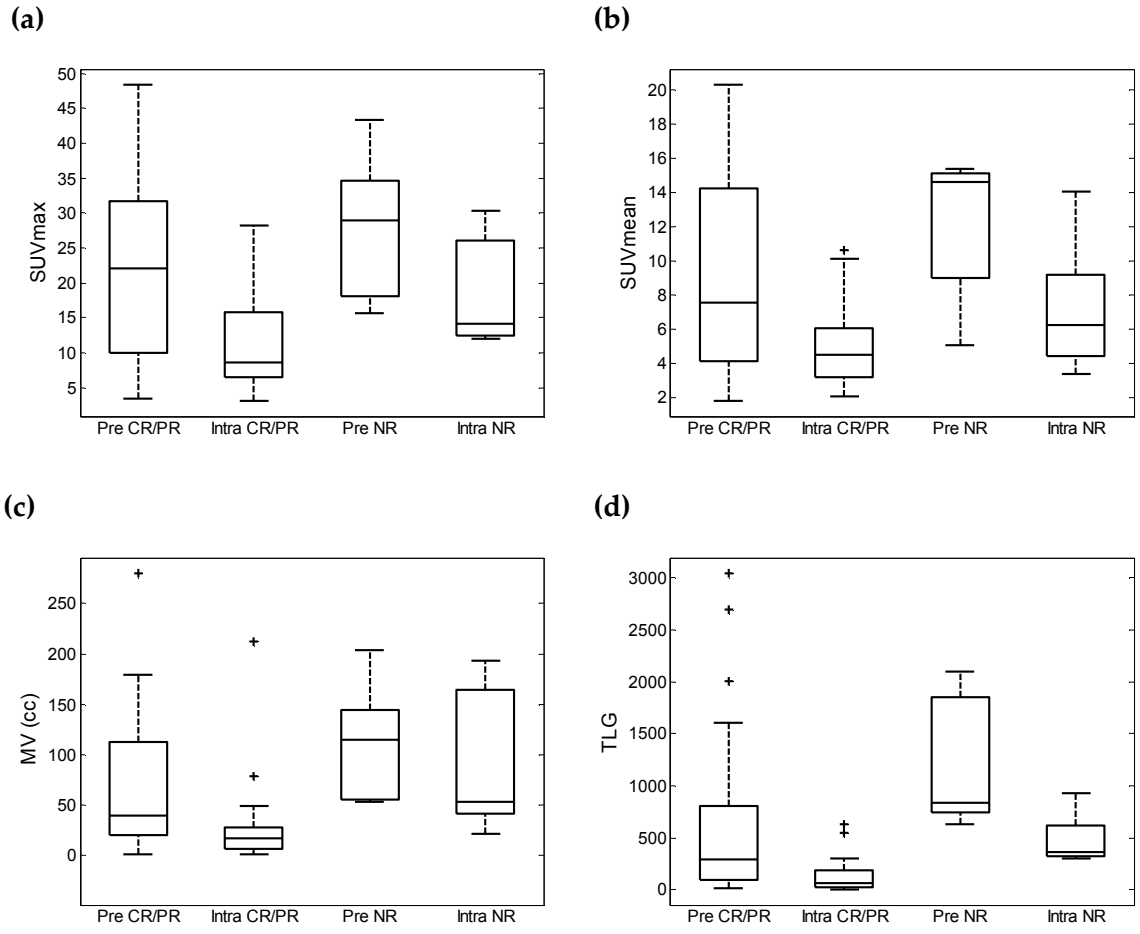


Figure 30: Comparison of baseline and intra-treatment 1st order features split between CR/PR and NR.

6.4.2 Discussion of Temporal Changes in 1st, 2nd, and High Order Features

As there were only 6 NR, no statistics were performed to quantify the changes in the responding and non-responding individual groups. The differences between responders and non-responders, however, can be clearly visualized. In general, texture features of responders initially correlated with positive prognosis and experienced larger changes during therapy, indicating greater tumor response. Texture features of

non-responding tumors initially correlated with poorer prognosis and in several cases, experienced less change in texture than the responding tumors. The changes in texture features correlate with SUV, MV, and TLG. As expected, responding tumors have a lower baseline and intra-treatment SUV than non-responders. Additionally, non-responding tumors are generally larger than responding tumors, and have higher TLG. In both responders and non-responders, tumors heterogeneity tends to decrease with treatment corresponding with a decrease in size and metabolic activity.

6.5 Characterization of Gynecological Tumors

It has been mentioned that tumors presenting areas of low signal may be caused by necrotic or hypoxic regions and are often resistant to radiation therapy. Of the 6 non-responding primary tumors, 4 presented signs of low uptake.

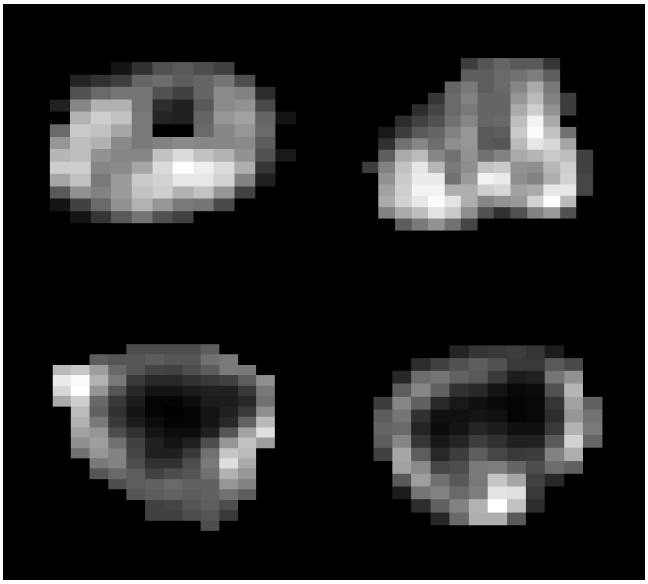


Figure 31: Primary PET GTVs of 4 non-responding tumors presenting signs of necrosis or hypoxia and corresponding features.

Texture analysis identifies physiological features of tumors that correlate with tumor response. Features such as *entropy, ZSNU,* and *complexity* are high in non-responders. These may be possible markers identifying areas as hypoxia and necrosis; attributes that correlate with negative outcomes in radiation therapy. Features such as *energy* and *max probability* correlating with responding tumors may identify homogeneity within a tumor.

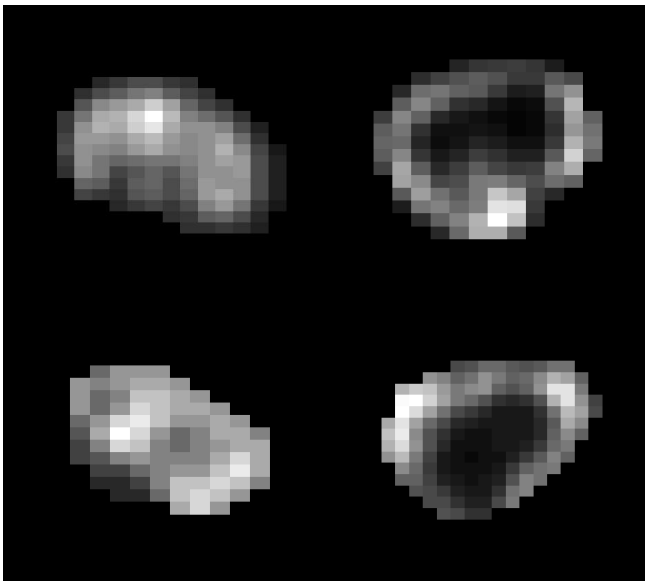


Figure 32: Examination of features associated with the images of a (a) baseline and (c) corresponding intra-treatment CR GTV compared to a (b) baseline and (d) corresponding intra-treatment NR GTV.

Our results indicate that texture analysis reflects the changing tumor physiology. While both responding and non-responding tumors experience a shift towards increased homogeneity, responding tumors experience a larger increase than non-responding tumors. Figure 32 depicts the tumor of a CR and NR from baseline to intra-treatment. Entropy of the CR decreases by 23% with therapy while entropy of the NR decreases by

only 2.4%. Interestingly, SUV_{max} and SUV_{mean} are higher in the CR than the NR while the percent decrease in SUV_{max} is smaller in the CR.

This comparison of a CR and NR GTV during therapy illustrates how texture analysis characterizes patterns correlating with outcome. We can see the increase in homogeneity of the CR while the NR remains heterogeneous and hypoxic even though the SUV values are relatively low. Imaging biomarkers are only effective if they can be associated with physiology. Our results indicate that texture analysis shows strong potential as an imaging biomarker capable of predicting tumor response to radiation therapy based on the underlying tissue characteristics.

7. Variability of Texture Features due to Reconstruction

Texture analysis examines the relationships between pixel intensities and is therefore sensitive to factors such as image noise. PET reconstruction methods effect image contrast and resolution and have been shown to change the outcome of texture analysis. Galavis et al. investigated the effect of various OSEM reconstruction parameters on texture features, but did not examine whether or not the variability due to reconstruction would prevent certain features from being meaningful. For example, if the variability among a given feature is greater than the difference between in that feature baseline and intra-treatment scans, it may cast doubt on its ability to reliably classify change during treatment.

The variability of texture features due to different reconstruction parameters in the TrueX + TOF algorithm from a Biograph mCT (Siemens) was examined. Baseline and intra-treatment PET scans from a subset of 7 patients were generated, varying the number of iterations and filter size. The number of subsets was maintained at 21.

Table 6: Parameters of the clinical and reconstructed PET image sets.

Patient Set	Iteration	Filter (Gaussian)
Clinical	2	2 mm
1	2	3 mm
2	2	4 mm
3	3	2 mm
4	3	3 mm
5	4	2 mm
6	4	3 mm

The primary GTV on the clinical baseline and intra-treatment PET set were contoured using PET Edge and transferred to each reconstructed set. Texture features were extracted from each reconstructed set and compared with the texture features from the clinical PET. The percent difference between the features from each baseline and intra-treatment reconstructed and clinical PET GTV was determined according to Equation 18.

$$\%\ Difference = \ 100 \times \frac{|X - X_{Clinical}|}{X_{Clinical}} \tag{18}$$

The mean percent differences for each feature were plotted in order to present an approximate range of variability. The features from the clinical and reconstructed

baseline and intra-treatment scans were then compared using the Wilcoxon signed-rank test to determine features that showed a statistically significant change ($p < 0.05$) during therapy. The baseline and intra-treatment reconstructed values for each feature were normalized to the baseline clinical value in order to limit variations to reconstruction effects. These results were compared with the statistics of all 29 patients. Features that showed significant changes in both groups may be considered as independent of reconstruction parameters.

7.1 Results of Varying Reconstruction Parameters.

Figure 33 presents an example of baseline clinical and reconstructed *entropy*. *Entropy* increased with filter size and iteration number.

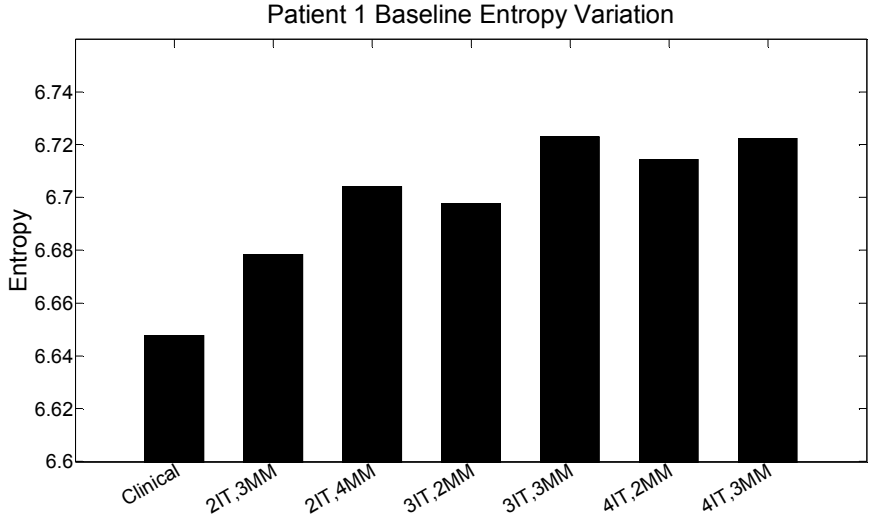


Figure 33: *Entropy* for clinical and 6 reconstructed PET image sets. Note that the range of variability is only 1.05%.

The percent ranges for first, second, and higher order features show a pattern of increased variability in intra-treatment PET scans.

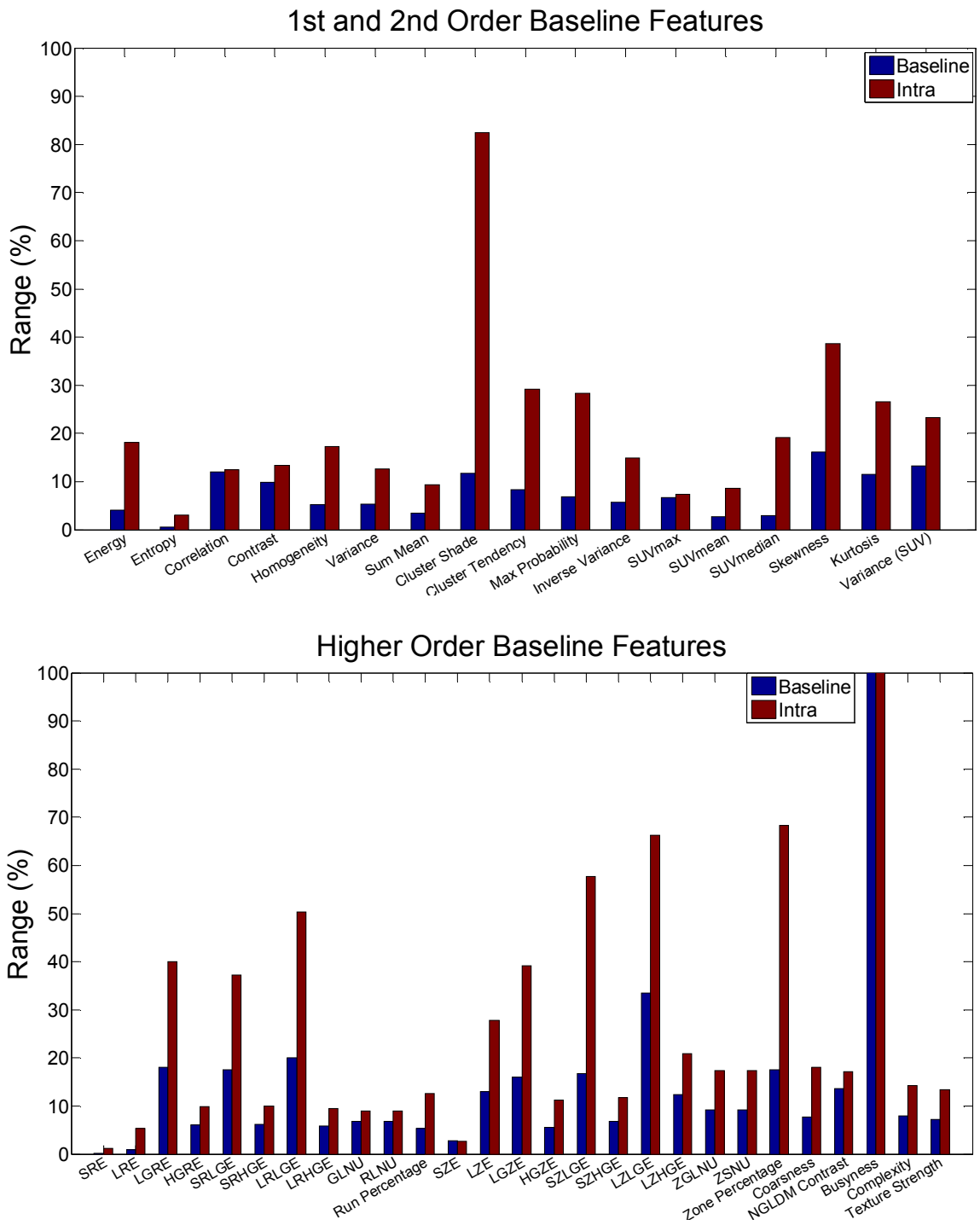


Figure 34: 1st, 2nd, and high order percent ranges for reconstructed baseline and intra-treatment PET scans.

As seen in Figure 34, the majority of baseline features remain under 10% mean range. Intra-treatment variability increased to over 10% in the majority of features. *Entropy*, *SRE, LRE*, and *SZE* were the least affected by varying reconstruction parameters. Variability in busyness was greater than 100% in the baseline scan and greater than 1000% in the intra-treatment scan. The other NGLDM experienced relatively low variability however.

Performing the Wilcoxon signed-rank test on all features yielded 25 texture features with $p < 0.05$. Comparing these features against the 14 texture features (Table 5) with $p > 0.05$ in the group of all 28 patients, 2 were showed no significant differences in the group of reconstructed features: *run percentage* and *zone percentage*. With the exception of *run percentage* and *zone percentage*, the features presented in Table 5 should be considered robust and independent of variability due to reconstruction parameters. Future work will include reconstructing more PET image sets and determining if there is any change in the predictive power of texture features with varying reconstruction parameters.

8. Conclusions

8.1 Summary and Discussion

The purpose of this thesis was to develop a method of predicting tumor response using baseline PET images of gynecological primary tumors as well as assess the changes in tumors during radiation therapy. Four gray level texture algorithms were adopted: the co-occurrence matrix, the run length matrix, the size zone matrix, and the neighborhood difference matrix. Second and high order features extracted from these matrices represent local and regional descriptors of tumor texture. Additionally, six 1st order features were assessed: SUV_{max}, SUV_{mean}, skewness, kurtosis, MV, and TLG. Duke FIRE, a MATLAB GUI, was developed and implemented in the clinical environment to expedite the extraction of 1st, 2nd, and high order features from PET images.

A total of 29 primary tumors were analyzed in this thesis, 28 of which had outcomes assessed by post-treatment imaging. Features extracted from baseline primary GTVs contoured by a semi-automatic gradient-based method, PET Edge, and by 40% SUV_{max} threshold allowed us to determine the prognostic capabilities of texture analysis compared to SUV metrics. Our results indicate that texture features provide a robust method of predicting tumor response. Additionally, we determined that PET Edge provides a more significant model for classifying tumors as responders or non-responders. The temporal changes of tumors during therapy were assessed using 1st, 2nd, and high order features as well. Texture features showed a strong correlation with

SUV metrics, indicating that tumors become increasingly more homogenous during radiation therapy. Images of non-responding tumors as well as a comparison between baseline and intra-treatment responding and non-responding tumors allows us to better understand the physiology characterized by texture analysis. Finally, variability in texture features due to reconstruction was explored in a subset of patients. The majority of features showing statistical changes during therapy are shown to be independent of reconstruction parameters.

Texture analysis shows exciting potential in the prediction of tumor response in radiation therapy. In the case of gynecological cancers, this study is the first to explore all 4 texture matrices and 1st order features for both outcome prediction and assessment.

8.2 Future Work

Using Duke FIRE, texture analysis can be easily extended to other disease sites for ongoing projects. Additionally, Duke FIRE has been extended to analyze MR ADC images and will soon allow for coronal ktrans image analysis. Between PET and MR features including texture analysis, we hope to establish a composite metric that will be robust in predicting tumor response from baseline images.

Appendix A

Appendix A presents each feature derived from the texture matrices as well as SUV intensity histogram.

Table 7: Formulations of 2nd order, local features derived from the GLCM

Feature	Equation	Definition
Energy	$\sum_i \sum_j P(i,j)^2$	Measures repeated pixel pairs.
Entropy	$\sum_i \sum_j P(i,j) \log P(i,j)$	Measures the randomness of the distribution.
Correlation	$\frac{\sum_i \sum_j (ij) P(i,j)^2 - \mu_x \mu_y}{\sigma_x \sigma_y}$	Increases with correlation between pixel pairs.
Contrast	$\sum_i \sum_j (i-j)^2 P(i,j)$	Measures similarity of pixel pairs.
Variance	$\frac{1}{2} \sum_i \sum_j (i-\mu_x)^2 P(i,j) + (i-\mu_y)^2 P(i,j)$	Measures the spread in pixel intensities.
Sum Average	$\frac{1}{2} \sum_i \sum_j iP(i,j) + jP(i,j)$	Average of pixel intensities.
Cluster Shade	$\sum_i \sum_j (i+j-\mu_x-\mu_y)^3 P(i,j)$	Measures the skewness of the GLCM. Increases with asymmetry.
Cluster Tendency	$\sum_i \sum_j (i+j-\mu_x-\mu_y)^4 P(i,j)$	Measures variations in pixel intensities.
Max Probability	$Max\ P(i,j)$	Reflects the highest number of occurrences of a pixel pair.
Inverse Variance	$\frac{\sum_i \sum_j P(i,j)}{(i-j)^2}$	Also called Inverse Difference Moment, similar to homogeity.
Homogeneity (From Haralick)	$\frac{\sum_i \sum_j P(i,j)}{1+(i-j)^2}$	Measures uniformity of image.
Homogeneity (From Tixier)	$\frac{\sum_i \sum_j P(i,j)}{1+\lvert i-j \rvert}$	Slightly different version of Haralick's formulation.

Table 8: Formulations of high order, regional features derived from the GLRLM

Feature	Equation	Definition
Short Run Emphasis	$\frac{1}{N_r}\sum_i\sum_j \frac{P(i,j)}{j^2}$	Measures the occurrence of short runs.
Long Run Emphasis	$\frac{1}{N_r}\sum_i\sum_j P(i,j)\times j^2$	Measures the occurrence of long runs.
Low Gray Level Run Emphasis	$\frac{1}{N_r}\sum_i\sum_j \frac{P(i,j)}{i^2}$	Measures the distribution of pixels weighted towards low intensity.
High Gray Level Run Emphasis	$\frac{1}{N_r}\sum_i\sum_j P(i,j)\times i^2$	Measures the distribution of pixels weighted towards high intensity.
Short Run Low Gray Level Emphasis	$\frac{1}{N_r}\sum_i\sum_j \frac{P(i,j)}{i^2\times j^2}$	Measures the occurrence of short runs weighted towards low intensity.
Long Run Low Gray Level Emphasis	$\frac{1}{N_r}\sum_i\sum_j \frac{P(i,j)\times i^2}{j^2}$	Measures the occurrence of long runs weighted towards low intensity.
Short Run High Gray Level Emphasis	$\frac{1}{N_r}\sum_i\sum_j \frac{P(i,j)\times j^2}{i^2}$	Measures the occurrence of short runs weighted towards high intensity.
Long Run High Gray Level Emphasis	$\frac{1}{N_r}\sum_i\sum_j P(i,j)\times i^2\times j^2$	Measures the occurrence of long runs weighted towards high intensity.
Run Gray Level Non-Uniformity	$\frac{1}{N_r}\sum_{i=1}^{M}\{\sum_{j=1}^{N} P(i,j)\}^2$	Measures the similarity of pixel intensities.
Run Length Non-Uniformity	$\frac{1}{N_r}\sum_{j=1}^{M}\{\sum_{i=1}^{N} P(i,j)\}^2$	Measures the similarity of lengths of pixels.
Run Percentage	$\frac{N_r}{P(i,j)\times j}$	Measures the percentage of runs of a given length. This value is 1 if the every run is of length 1.

Table 9: Formulations of high order, regional features derived from the GLSZM

Feature	Equation	Definition
Small Zone Emphasis	$\frac{1}{N_r}\sum_i\sum_j \frac{P(i,j)}{j^2}$	Measures the occurrence of small zones.
Large Zone Emphasis	$\frac{1}{N_r}\sum_i\sum_j P(i,j)\times j^2$	Measures the occurrence of large zones.
Low Gray Level Zone Emphasis	$\frac{1}{N_r}\sum_i\sum_j \frac{P(i,j)}{i^2}$	Measures the distribution of pixels weighted towards low intensity.
High Gray Level Zone Emphasis	$\frac{1}{N_r}\sum_i\sum_j P(i,j)\times i^2$	Measures the distribution of pixels weighted towards high intensity.
Small Zone Low Gray Level Emphasis	$\frac{1}{N_r}\sum_i\sum_j \frac{P(i,j)}{i^2\times j^2}$	Measures the occurrence of small zones weighted towards low intensity.
Large Zone Low Gray Level Emphasis	$\frac{1}{N_r}\sum_i\sum_j \frac{P(i,j)\times i^2}{j^2}$	Measures the occurrence of large zones weighted towards low intensity.
Small Zone High Gray Level Emphasis	$\frac{1}{N_r}\sum_i\sum_j \frac{P(i,j)\times j^2}{i^2}$	Measures the occurrence of small zones weighted towards high intensity.
Large Zone High Gray Level Emphasis	$\frac{1}{N_r}\sum_i\sum_j P(i,j)\times i^2\times j^2$	Measures the occurrence of large zones weighted towards high intensity.
Zone Gray Level Non-Uniformity	$\frac{1}{N_r}\sum_{i=1}^{M}\{\sum_{j=1}^{N} P(i,j)\}^2$	Measures the similarity of pixel intensities.
Zone Size Non-Uniformity	$\frac{1}{N_r}\sum_{j=1}^{M}\{\sum_{i=1}^{N} P(i,j)\}^2$	Measures the similarity of zones of pixels.
Zone Percentage	$\frac{N_r}{P(i,j)\times j}$	Measures the percentage of zones of a given size. This value is 1 if the every zone is of size 1.

Table 10: Formulations of high order, local features derived from the NGLDM

Feature	Equation	Definition
Coarseness	$\left[\varepsilon + \sum_{i=0}^{G_h} p_i s(i)\right]$ $p_i = \frac{N_i}{n^3}, n = N - 2d$	A summation of the intensity differences throughout the image weighted by the probability of occurrence.
Contrast	$\left[\frac{1}{N_g(N_g-1)} \sum_{i=0}^{G_h}\sum_{j=0}^{G_h} p_i\, p_j\, (i-j)^2\right]\left[\frac{1}{n^3}\sum_{i=0}^{G_h} s(i)\right]$ $p_i \neq 0, p_j \neq 0$	Measures the difference between neighboring regions of pixel intensities.
Busyness	$\left[\sum_{i=0}^{G_h} p_i s(i)\right] / \left[\sum_{i=0}^{G_h}\sum_{j=0}^{G_h} (ip_i - jp_j)\right]$ $p_i \neq 0, p_j \neq 0$	Measures the rate of change between neighborhood intensities weighted by the difference in intensities.
Complexity	$\left\{\sum_{i=0}^{G_h}\sum_{j=0}^{G_h} \frac{\lvert i-j\rvert}{n^3(p_i+p_j)}\right\}\{p_i s(i) + p_j s(j)\}$ $p_i \neq 0, p_j \neq 0$	Measures the uniformity of patterns versus rate of change in an image.
Texture Strength	$\left[\sum_{i=0}^{G_h}\sum_{j=0}^{G_h} (p_i + p_j)\,(i-j)^2\right] / \left[\varepsilon + \sum_{i=0}^{G_h} s(i)\right]$ $p_i \neq 0, p_j \neq 0$	Measures the differences in intensities divided by the size of the areas.

Table 11: Formulations of 1st order, global features

Feature	Equation	Definition
Skewness	$\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\sigma^3}$	Measures the asymmetry about the sample mean.
Kurtosis	$\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\sigma^4}$	Measures the peak compared to a normal distribution. (normal = 3)
Variance	$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$	Measures the spread of the data.

References

[1] L.A. Dawson, C. Menard, Imaging in radiation oncology: a perspective, The oncologist, 15 (2010) 338-349.

[2] R. Lumia, R.M. Haralick, O. Zuniga, L. Shapiro, T.-C. Pong, F.-P. Wang, Texture analysis of aerial photographs, Pattern Recogn, 16 (1983) 39-46.

[3] J.S. Weszka, A. Rosenfeld, An application of texture analysis to materials inspection, Pattern Recogn, 8 (1976) 195-200.

[4] O. Guntinas-Lichius, E. Schulte, E. Stennert, W.F. Neiss, The use of texture analysis to study the time course of chromatolysis, Journal of Neuroscience Methods, 78 (1997) 1-6.

[5] X. Yang, S. Tridandapani, J.J. Beitler, D.S. Yu, E.J. Yoshida, W.J. Curran, T. Liu, Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: an in vivo study of late toxicity, Medical physics, 39 (2012) 5732-5739.

[6] S. Chicklore, V. Goh, M. Siddique, A. Roy, P.K. Marsden, G.J.R. Cook, Quantifying tumour heterogeneity in F-18-FDG PET/CT imaging by texture analysis, European journal of nuclear medicine and molecular imaging, 40 (2013) 133-140.

[7] N.M. Cheng, Y.H.D. Fang, T.C. Yen, The promise and limits of PET texture analysis, Ann Nucl Med, 27 (2013) 867-869.

[8] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012, International journal of cancer. Journal international du cancer, 136 (2015) E359-386.

[9] V.T. DeVita, T.S. Lawrence, S.A. Rosenberg, DeVita, Hellman, and Rosenberg's cancer : principles & practice of oncology, 9th ed., Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, 2011.

[10] N. Sakuragi, C. Satoh, N. Takeda, H. Hareyama, M. Takeda, R. Yamamoto, T. Fujimoto, M. Oikawa, T. Fujino, S. Fujimoto, Incidence and distribution pattern of pelvic and paraaortic lymph node metastasis in patients with Stages IB, IIA, and IIB cervical carcinoma treated with radical hysterectomy, Cancer, 85 (1999) 1547-1554.

[11] W.T. Creasman, C.P. Morrow, B.N. Bundy, H.D. Homesley, J.E. Graham, P.B. Heller, Surgical pathologic spread patterns of endometrial cancer. A Gynecologic Oncology Group Study, Cancer, 60 (1987) 2035-2041.

[12] J.A. Green, J.M. Kirwan, J.F. Tierney, P. Symonds, L. Fresco, M. Collingwood, C.J. Williams, Survival and recurrence after concomitant chemotherapy and radiotherapy for cancer of the uterine cervix: a systematic review and meta-analysis, Lancet, 358 (2001) 781-786.

[13] L. Portelance, K.S. Chao, P.W. Grigsby, H. Bennet, D. Low, Intensity-modulated radiation therapy (IMRT) reduces small bowel, rectum, and bladder doses in patients with cervical cancer receiving pelvic and para-aortic irradiation, International journal of radiation oncology, biology, physics, 51 (2001) 261-266.

[14] A.J. Mundt, A.E. Lujan, J. Rotmensch, S.E. Waggoner, S.D. Yamada, G. Fleming, J.C. Roeske, Intensity-modulated whole pelvic radiotherapy in women with gynecologic malignancies, International journal of radiation oncology, biology, physics, 52 (2002) 1330-1337.

[15] P.W. Grigsby, The prognostic value of PET and PET/CT in cervical cancer, Cancer imaging : the official publication of the International Cancer Imaging Society, 8 (2008) 146-155.

[16] J.S. Karp, S. Surti, M.E. Daube-Witherspoon, G. Muehllehner, Benefit of time-of-flight in PET: experimental and clinical results, J Nucl Med, 49 (2008) 462-470.

[17] K.A. Higgins, J.K. Hoang, M.C. Roach, J. Chino, D.S. Yoo, T.G. Turkington, D.M. Brizel, Analysis of pretreatment FDG-PET SUV parameters in head-and-neck cancer: tumor SUVmean has superior prognostic value, International journal of radiation oncology, biology, physics, 82 (2012) 548-553.

[18] D. Oh, J.E. Lee, S.J. Huh, W. Park, H. Nam, J.Y. Choi, B.T. Kim, Prognostic significance of tumor response as assessed by sequential 18F-fluorodeoxyglucose-positron emission tomography/computed tomography during concurrent chemoradiation therapy for cervical cancer, International journal of radiation oncology, biology, physics, 87 (2013) 549-554.

[19] N.P. Rizk, L. Tang, P.S. Adusumilli, M.S. Bains, T.J. Akhurst, D. Ilson, K. Goodman, V.W. Rusch, Predictive value of initial PET-SUVmax in patients with locally advanced esophageal and gastroesophageal junction adenocarcinoma, Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer, 4 (2009) 875-879.

[20] Y. Zaizen, K. Azuma, S. Kurata, E. Sadashima, S. Hattori, T. Sasada, Y. Imamura, H. Kaida, A. Kawahara, T. Kinoshita, M. Ishibashi, T. Hoshino, Prognostic significance of total lesion glycolysis in patients with advanced non-small cell lung cancer receiving chemotherapy, European journal of radiology, 81 (2012) 4179-4184.

[21] E.H. Dibble, A.C. Alvarez, M.T. Truong, G. Mercier, E.F. Cook, R.M. Subramaniam, 18F-FDG metabolic tumor volume and total glycolytic activity of oral cavity and oropharyngeal squamous cell cancer: adding value to clinical staging, J Nucl Med, 53 (2012) 709-715.

[22] M. Hatt, C. Cheze le Rest, A. Turzo, C. Roux, D. Visvikis, A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET, IEEE transactions on medical imaging, 28 (2009) 881-893.

[23] X. Geets, J.A. Lee, A. Bol, M. Lonneux, V. Gregoire, A gradient-based method for segmenting FDG-PET images: methodology and validation, European journal of nuclear medicine and molecular imaging, 34 (2007) 1427-1438.

[24] M. Wanet, J.A. Lee, B. Weynand, M. De Bast, A. Poncelet, V. Lacroix, E. Coche, V. Gregoire, X. Geets, Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: a comparison with threshold-based approaches, CT and surgical specimens, Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology, 98 (2011) 117-125.

[25] R.J. McGurk, J. Bowsher, J.A. Lee, S.K. Das, Combining multiple FDG-PET radiotherapy target segmentation methods to reduce the effect of variable performance of individual segmentation methods, Medical physics, 40 (2013) 042501.

[26] G. Zhang, D. Han, C. Ma, J. Lu, T. Sun, T. Liu, J. Zhu, J. Zhou, Y. Yin, Gradient-based delineation of the primary GTV on FLT PET in squamous cell cancer of the thoracic esophagus and impact on radiotherapy planning, Radiat Oncol, 10 (2015) 11.

[27] H. Zaidi, M. Abdoli, C.L. Fuentes, I.M. El Naqa, Comparative methods for PET image segmentation in pharyngolaryngeal squamous cell carcinoma, European journal of nuclear medicine and molecular imaging, 39 (2012) 881-891.

[28] K.H. Zou, S.K. Warfield, A. Bharatha, C.M. Tempany, M.R. Kaus, S.J. Haker, W.M. Wells, 3rd, F.A. Jolesz, R. Kikinis, Statistical validation of image segmentation quality based on a spatial overlap index, Academic radiology, 11 (2004) 178-189.

[29] R.M. Haralick, Shanmuga.K, I. Dinstein, Textural Features for Image Classification, Ieee T Syst Man Cyb, Smc3 (1973) 610-621.

[30] A.S. Kurani, D.-H. Xu, J. Furst, D.S. Raicu, Co-occurrence matrices for volumetric data, in: 7th IASTED International Conference on Computer Graphics and Imaging, Kauai, USA, 2004, pp. 447-452.

[31] M.M. Galloway, Texture analysis using gray level run lengths, Computer graphics and image processing, 4 (1975) 172-179.

[32] D.-H. Xu, A.S. Kurani, J.D. Furst, D.S. Raicu, Run-length encoding for volumetric texture, Heart, 27 (2004) 25.

[33] G. Thibault, B. Fertil, C. Navarro, S. Pereira, P. Cau, N. Levy, J. Sequeira, J. Mari, Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification, (2009).

[34] M. Amadasun, R. King, Textural Features Corresponding to Textural Properties, Ieee T Syst Man Cyb, 19 (1989) 1264-1274.

[35] X. Dong, L. Xing, P. Wu, Z. Fu, H. Wan, D. Li, Y. Yin, X. Sun, J. Yu, Three-dimensional positron emission tomography image texture analysis of esophageal squamous cell carcinoma: relationship between tumor 18F-fluorodeoxyglucose uptake heterogeneity, maximum standardized uptake value, and tumor stage, Nuclear medicine communications, 34 (2013) 40-46.

[36] M. Soussan, F. Orlhac, M. Boubaya, L. Zelek, M. Ziol, V. Eder, I. Buvat, Relationship between Tumor Heterogeneity Measured on FDG-PET/CT and Pathological Prognostic Factors in Invasive Breast Cancer, Plos One, 9 (2014).

[37] J.M. Willaime, F.E. Turkheimer, L.M. Kenny, E.O. Aboagye, Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of 18F fluorothymidine-positron emission tomography, Physics in medicine and biology, 58 (2013) 187-203.

[38] F. Tixier, C.C. Le Rest, M. Hatt, N. Albarghach, O. Pradier, J.P. Metges, L. Corcos, D. Visvikis, Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer, J Nucl Med, 52 (2011) 369-378.

[39] U. Nestle, S. Kremp, A. Schaefer-Schuler, C. Sebastian-Welsch, D. Hellwig, C. Rube, C.M. Kirsch, Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer, J Nucl Med, 46 (2005) 1342-1348.

[40] F. Orlhac, M. Soussan, J.A. Maisonobe, C.A. Garcia, B. Vanderlinden, I. Buvat, Tumor Texture Analysis in F-18-FDG PET: Relationships Between Texture Parameters, Histogram Indices, Standardized Uptake Values, Metabolic Volumes, and Total Lesion Glycolysis, J Nucl Med, 55 (2014) 414-422.

[41] I. El Naqa, P. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, S. Chaudhari, D. Yang, M. Schmitt, R. Laforest, W. Thorstad, J.O. Deasy, Exploring feature-based approaches in PET images for predicting cancer treatment outcomes, Pattern Recognit, 42 (2009) 1162-1171.

[42] F. Yang, M.A. Thomas, F. Dehdashti, P.W. Grigsby, Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer, European journal of nuclear medicine and molecular imaging, 40 (2013) 716-727.

[43] E.A. Eisenhauer, P. Therasse, J. Bogaerts, L.H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, J. Verweij, New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1), European journal of cancer, 45 (2009) 228-247.

[44] G. Guyatt, R. Jaeschke, N. Heddle, D. Cook, H. Shannon, S. Walter, Basic statistics for clinicians: 2. Interpreting study results: confidence intervals, CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 152 (1995) 169-173.

[45] M.H. Zweig, G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, Clinical chemistry, 39 (1993) 561-577.

[46] D.W. Hosmer Jr, S. Lemeshow, Applied logistic regression, John Wiley & Sons, 2004.

[47] P. Vaupel, A. Mayer, Hypoxia in cancer: significance and impact on clinical outcome, Cancer metastasis reviews, 26 (2007) 225-239.

[48] M. Hockel, K. Schlenger, S. Hockel, B. Aral, U. Schaffer, P. Vaupel, Tumor hypoxia in pelvic recurrences of cervical cancer, International journal of cancer. Journal international du cancer, 79 (1998) 365-369.