UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

QUANTITATIVE CT IMAGE FEATURE ANALYSIS FOR PREDICTING TUMOR RESPONSE TO CHEMOTHERAPY AT EARLY STAGE

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

GOPICHANDH DANALA
Norman, Oklahoma
2016

QUANTITATIVE CT IMAGE FEATURE ANALYSIS FOR PREDICTING TUMOR RESPONSE TO CHEMOTHERAPY AT EARLY STAGE

A THESIS APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY

_____
Dr. Yuchen Qiu, Chair

_____
Dr. Bin Zheng

_____
Dr.Liangzhong Xiang

Acknowledgements

The author would like to gratefully acknowledge his thesis advisor Dr. Yuchen Qiu. Dr. Qiu offered the author this opportunity to perform his study, and mentored him to accomplish his master degree thesis.

The author would also like to express his gratitude to the radiologists and oncologists in Health Sciences Center, University of Oklahoma (OUHSC): Dr. Kathleen M. Moore, Dr. Katherine M. Moxley, Dr. Camille C. Gunderson and Dr. Theresa C. Thai, for their clinical support on his study.

The author would also like to thank all his committee members: Dr. Bin Zheng and Dr. Liangzhong Xiang, for their efforts of serving in the author's M.S. program committee.

Finally, author expresses his deep gratitude to his family, for their endless support both inside and outside the research.

Table of Contents

List of Tables

List of Figures

Abstract

In gynecologic oncology, ovarian cancer is the second leading cancer with highest mortality rate. Since most of ovarian cancer patients are diagnosed at advanced stage occurred with metastatic tumors, chemotherapy is a necessary treating procedure after the aggressive surgery which removes the patient's primary ovary tumor (Primary cytoreduction). However, currently, no method is able to effectively and efficiently predict the tumor response to the chemotherapy at early stage (i.e. 4-6 weeks after the treatment). This study aims to investigate whether using quantitative image features computed from CT images enables to more accurately predict response of ovarian cancer patients to chemotherapy. During the experiment, we retrospectively assembled a dataset involving 91 patients. Each patient had two sets of pre-and post-therapy (4-6 weeks follow up) CT images. A computer-aided detection scheme was then developed, which is able segment the metastatic tumors and computed image features. Next, we built two initial feature pools using image features computed from pre-therapy CT images only and image feature difference computed from both pre- and post-therapy images. The predicting performance of each feature was evaluated using the area under ROC curve, which is based on the criteria 6-month progression-free survival (PFS). Among these features, the optimal feature cluster was determined and an equal-weighted fusion method was used to generate a new image marker to predict PFS of the patients. The results indicate that the highest single feature AUC values are achieved as 0.6842 ± 0.0557 and 0.7705 ± 0.0495 respectively, which are computed from pre-therapy CT images only and both pre- and post-therapy CT images. When applying fusion-based image markers, AUC values significantly increased to 0.8103 ± 0.0447 and 0.8292 ± 0.0431 ($p < 0.05$),

respectively. This study demonstrated that it is feasible to predict patients' early stage response to chemotherapy using quantitative image features computed from pre-therapy CT images. However, we can significantly improve the prediction performance when adding information from the 4-6 week follow up CT images.

Chapter 1: Introduction

Ovarian Cancer

Ovarian cancer is the second most common cancer in gynecologic oncology. In 2016, approximately 22,280 newly diagnosed cases and 14,240 deaths are estimated in United States [1]. Given that early symptoms originating from ovaries are hard to be detected and even more likely to be confused for something far less serious, routine pelvic examinations are not able to effectively detect ovarian cancer. As a result, approximately 70% of ovarian cancer patients' are diagnosed at an advanced stage [2] with metastatic tumors, and ovarian cancer is the leading cause of death among all the gynecologic cancers. As indicated in Figure 1, the five-year survival rates for stage III and IV epithelial ovarian cancer patients are only 36% and 17% respectively [3]. For the advanced stage ovarian cancer patients, an aggressive surgery is first operated to remove the primary ovarian tumors (Primary cytoreduction) and chemotherapy is then followed to control the metastatic tumors. Since ovarian cancer is highly heterogeneous, and the metastasized tumors are typically P53 driven and genetically instable [4], the chemotherapy response varies significantly among the patients. Therefore, a large number of clinical trials have been performed to develop and test the efficacy of new biologically targeted agents, drugs and/or chemotherapeutic procedures. However, one of the major challenges in these clinical trials is the limited ability to accurately categorize patients to find a certain group of patients which is more likely to respond to target treatment. In addition, most of the chemotherapy drugs have significant side effects and high costs. As a result, there is a clinically imperative need for developing an early stage prognostic assessment method to

accurately categorize the patients into groups that are likely or not to benefit from the new therapies.

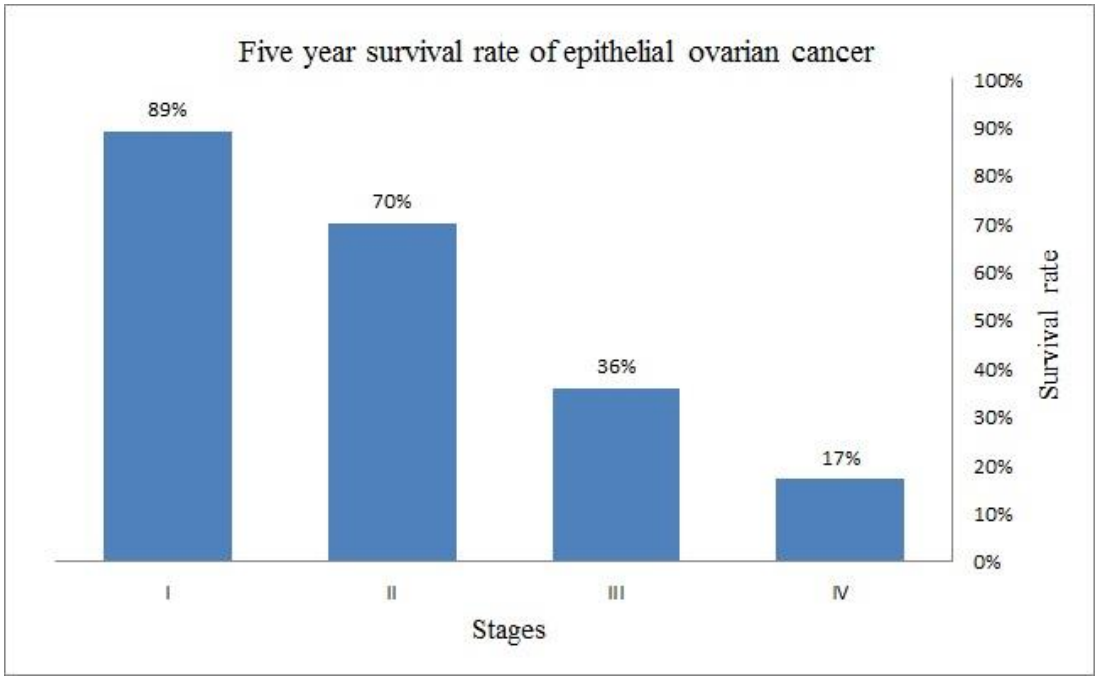


Figure 1: Five-year survival rate of epithelial ovarian cancer patients according to National Cancer Institute, SEER Data Base for the patients diagnosed from 1995 to 2007 [3]

Evaluating early response to chemotherapy drugs

Extensive researches have been conducted to develop novel clinical markers for prognostic assessment of tumor response to chemotherapy, which can be classified into two subgroups: genetic biomarkers and radiographic imaging. In the recent years, many genetic biomarkers have been discovered and investigated [4-6] to select optimal treatment methods [7-9]. However due to the complexity and heterogeneity of ovarian cancer, no existing biomarkers are able to accurately select treatment options, predict clinical benefit, and determine drug resistance to date. This could be attributed by the facts that the current biomarkers are (1) applicable only to a certain group of patients [8-

10], (2) high costs [11-12] and (3) lower specificity [13-15]. Hence, radiographic imaging analyses are critically important to assess the prognostic response of ovarian cancer treatment [5-6].

Radiographic imaging

Current imaging modalities in treating ovarian cancer include ultrasound, perfusion X-ray computed tomography (CT), magnetic resonance imaging (MRI) and Positron emission tomography (PET). Due to several benefits like wider availability, rapid acquisition of images, a wealth of clear and specific information, high diagnostic performance, and low operating cost, perfusion X-ray CT is considered to be the most popular technique currently used for prognostic assessment of ovarian cancer in clinical practice [7].

RECIST guidelines

Response Evaluation Criteria in Solid Tumors (RECIST) is the current clinical standard to assess patients' response to treatment [8-9] (i.e. tumors in cancer patients increase, stabilize or decrease). RECIST evaluation includes following steps: (1). Identify, record and measure all measurable lesions up to 2 per organ and a maximum of 5 in total for all organs. The target lesions are selected based on their size/ longest diameter (LD) and their suitability for further measurements. (2). Compute the sum of the LD's of all the target lesions for response assessment. The assessment can be classified into four groups: Complete Response (CR): Disappearance of all target lesions; Partial Response (PR): Decrease of at least 30% in the sum of LD's of all the target lesions;

Progressive Disease (PD): Increase of at least 20% in the sum of LD's of all the target lesions; Stable Disease (SD): Insignificant increase or decrease in the sum of LD's of target lesions to be classified as either PD or PR respectively.

Radiologist assessment

In order to assess the patients' response to new chemotherapy drugs or procedures, radiographic perfusion CT examinations were performed on each patient. This study includes one set of CT scan images taken pre-therapy and a multiple sets of CT scans taken during the period of post therapy. In order to assess and/or categorize patients' response, radiologists will examine and mark up to 5 metastatic tumors from the pre- and post- treatment CT images, using RECIST guidelines. However, these results are based on one-dimensional size and don't take changes in tumor volume [10] to consideration often leading to a low association between radiologist results and clinical outcome [11].

Computer Aided Detection (CAD) scheme

In order to improve the early prognostic assessment of patients' response and assist the radiologist in making optimal treatment decisions for ovarian cancer patients, a CAD scheme was developed to more accurately assess the tumor response to chemotherapy, using pre- and post- treatment CT images. The scheme firstly segments tumors marked by radiologist and then computes quantitative image features to detect and quantify tumor volume, shape, and density heterogeneity. The individual and combined features are finally used to predict 6-month progression free survival (PFS) of the patients, which is both recommended and approved criterion by Food and Drug Administration

(FDA) of the United States and European Medicines Agency (EMA) [12]. The overall objective of this study is to examine 1) whether quantitative image features computed from pre-therapy CT images offer information to predict patient response prior to chemotherapy and 2) whether quantitative image features computed from both pre- and post- therapy CT images offer significantly higher prediction capabilities to predict ovarian cancer patients response to chemotherapy than using traditional RECIST criteria.

Chapter 2: Materials

Image dataset

Under institutional review board (IRB) approved protocol, we retrospectively collected a dataset consisting of de-identified CT images acquired from 91 ovarian cancer patients. Each patient was diagnosed with recurrent, high-grade (serous, endometrioid or mixed) ovarian/ peritoneal /tubal carcinoma and underwent a clinical trial. The clinical trials involved 32 different new therapeutic drugs, each patient was given a mixture of 1/2/3 drugs which depends on her condition and doctor's treatment decision. Two sets of CT scans: pre-treatment and 4-6 weeks post-treatment were acquired for each patient. These CT images were obtained based on a pre-established CT scanning protocol in our medical center using either a 64-row detector CT machine (Light Speed VCT, GE Healthcare, Milwaukee, WI, USA) or a 16-row detector CT machine (Discovery 600, GE Healthcare). The X-ray power was operated at 120 kVp and 100-600mAs, depending upon the patient body size. A 100cc of contrast agent was injected into patient's body at a rate of 2-3cc/sec before CT examinations for better visualization of the tumor.

For each case, radiologist tracked the metastatic tumors in both pre- and post-treatment CT image, and the tumor size changes were estimated based on RECIST criteria. Since our study is retrospective, the 6 month PFS were also collected in the dataset for the performance assessment. In this dataset, 52 patients remained 6-month PFS (CR, PR, SD), while 39 did not maintain 6-month PFS to the treatment with signs of progressive disease (PD). Table 1 summarizes the detailed information about both the classes of patients

Table 1: Detailed information about patients separated by PFS

	PFS "Yes"	PFS "No"
Number of patients	52	39
Patient average age	66 ± 8	67 ± 9
Number of metastatic tumors	101	96
Average tumor diameter (before therapy)	27 mm	24mm

* The chemotherapy agent given varies among patients

Chapter 3: Quantitative image features analysis

Segmentation methods

In order to segment the tumors effectively and efficiently, we developed a segmentation scheme consisting of 8 different segmentation algorithms. This segmentation scheme can be categorized in to 3 sub-groups: 1) region growing methods, 2) canny operator based methods and 3) partial differential equation based methods. Figure 2 illustrates the flowchart for these three types of methods.

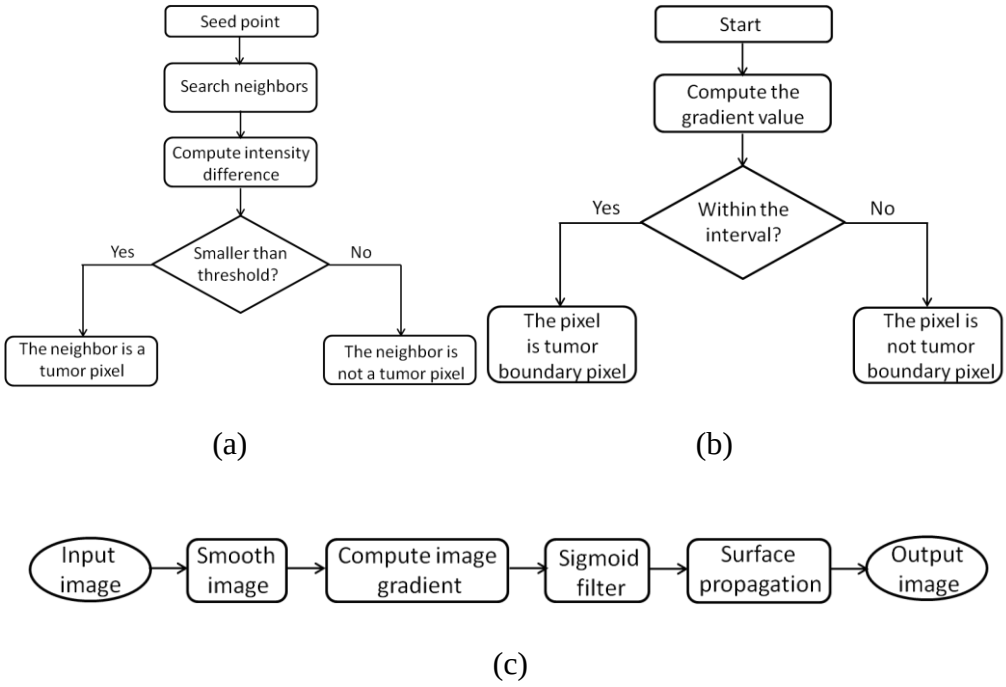


Figure 2: Flow chart of (a) region growing, (b) canny operator and (c) partial differential equation based segmentation algorithms

Region growing methods

This sub-group consists of two segmentation algorithms: connected and neighborhood region growing segmentation. Connected region growing is a modified form of regular region growing algorithm. For the connected region growing method, we initially select a seed point within the tumor and then evaluate whether its neighboring pixels are part of the tumor or not. If the intensity difference between neighbor pixel and the initial seed point is below the pre-determined threshold value, the neighbor pixel is considered as a part of tumor. This process is repeated for all the connected pixels until no neighboring pixel can satisfy the threshold condition. Neighborhood region growing is similar to the connected region growing method, and the only difference is that the pixel will be considered as part of the tumor if all its neighbors satisfy the intensity difference threshold.

Canny operator methods

This sub-group consists of two algorithms: Canny boundary extraction and multi-scale canny methods. Canny boundary extraction method includes three steps. First, A Gaussian filter is used to suppress the noise in the image. Next, a gradient intensity image is computed, for which hysteresis threshold is applied to optimize the edge detection. For this method, two thresholds (high and low thresholds) are used on the gradient intensity image, where high threshold filters the noise and maps the genuine edges in the image (T1), and low threshold is applied to track back the faint edges in T1 and make a continuous boundary image (T2). Finally, dilation is applied from the initial set seed point in all the direction until it reaches the traced continuous boundary to extract the tumor. Multi-scale canny method is a modified form of canny edge detection algorithm which

uses down sampled image with different scales. In canny edge detection, the resulting image can be considered as the convolution between the original image F and first derivative of the Gaussian kernel θ:

$$\frac{\partial F(x,y)\otimes\theta(x,y)}{\partial x} = F(x,y)\otimes\frac{\partial\theta(x,y)}{\partial x} = F(x,y)\otimes\psi^1(x,y) \quad (1)$$

$$\frac{\partial F(x,y)\otimes\theta(x,y)}{\partial y} = F(x,y)\otimes\frac{\partial\theta(x,y)}{\partial y} = F(x,y)\otimes\psi^2(x,y) \quad (2)$$

In case of multi-scale canny, we use dilated Gaussian kernel as follows:

$$\psi^1_{2^j} = \frac{1}{2^j}\psi^1\left(\frac{x}{2^j},\frac{y}{2^j}\right) \quad (3)$$

$$\psi^2_{2^j} = \frac{1}{2^j}\psi^2\left(\frac{x}{2^j},\frac{y}{2^j}\right) \quad (4)$$

Thus we compute the dilated derivative image at different scales:

$$D^1_{2^j}\left(\frac{\partial F(x,y)\otimes\theta(x,y)}{\partial x}\right) = F(x,y)\otimes\psi^1_{2^j} \quad (5)$$

$$D^2_{2^j}\left(\frac{\partial F(x,y)\otimes\theta(x,y)}{\partial y}\right) = F(x,y)\otimes\psi^2_{2^j} \quad (6)$$

In this study, the multi-scale edge detection is accomplished by applying the canny operator on the down sampled image with different scales. Given that most tumors are small, the scale number j is set to be 1 in this experiment. The multi-scale canny edge detecting operation is implemented by performing the canny operator on the down sampled images. After the operation, the extracted boundary is up sampled to the original scale.

Partial differential equation methods

The algorithms in this group are based on level set segmentation and consist of four algorithms: Fast marching, shape detection, geometric active contour and threshold level set methods. For level set methods, the boundary function $f(x, y)$ is defined as a curve when setting the three dimensional surface ψ to be a constant number C: $\psi(x, y) = C$. In most cases, the three dimensional initial surface is a paraboloid and its boundary curve is a circle. Next, the surface propagates based on the gradient values of the target image and its boundary changes accordingly. Surface propagation can be described using partial differential equation as below:

$$\frac{\partial \psi}{\partial t} + F|\nabla \psi| = 0 \tag{7}$$

$$\frac{\partial \psi}{\partial t} = -F_A|\nabla \psi| - F_G|\nabla \psi| - F_k k|\nabla \psi| \tag{8}$$

Where F_A is the advection speed given by the user, F_G is the local geometric speed term as follows:

$$F_G = -\frac{F_A}{(M_1 - M_2)}\{|\nabla G_\sigma * I(x, y)| - M_2\} \tag{9}$$

Where M_1, M_2 are the maximum and minimum of image gradients, ∇G_σ is the gradient Gaussian kernel, *I(x, y)* is the image. Accordingly, the total speed of the surface propagation will be slow at the high gradient image area and high at the low gradient area. Given that the image gradient will be maximized at the tumor boundary, the surface will stop or move very slowly at the boundary.

In fast marching, the image is initially pre-processed using three steps. First, an anisotropic filter is applied to smooth the image. Next, gradient image is computed. After that, the gradient image is passed to a sigmoid filter. Using the gradient image, the local

propagation speed F_G is estimated by formula (9). The surface propagation is finally determined by F_A, and F_G:

$$\frac{\partial \psi}{\partial t} = -F_A|\nabla\psi| - F_G|\nabla\psi| \quad (10)$$

In shape detection, the surface propagation uses all the three terms including mean curvature term (F_k). This term improves the boundary propagation by removing the noisy area in the target image.

$$\frac{\partial \psi}{\partial t} = -F_A|\nabla\psi| - F_G|\nabla\psi| - F_k k|\nabla\psi| \quad (11)$$

Where curvature term F_k is estimated as follows:

$$k = -\frac{\psi_{xx}\psi_y^2 - 2\psi_x\psi_x\psi_{xy} + \psi_{yy}\psi_x^2}{(\psi_x^2+\psi_y^2)^{3/2}} \quad (12)$$

In Geometric active contour, a second advection term is added to the surface propagation which is able to attract the computed contour to the target tumor boundary.

$$\frac{\partial \psi}{\partial t} = -F_A|\nabla\psi| - \alpha g(I)|\nabla\psi| - F_G|\nabla\psi| - F_k k|\nabla\psi| \quad (13)$$

For the last algorithm, threshold level set is a combination of both region growing and level set algorithms. A pre-determined threshold interval [LT, UT] is used to identify the tumor, while the smoothness of the propagation surface stops the leakage that usually happens in region growing algorithms.

The sample segmentation images of all the above discussed eight segmentation techniques are shown in the Figure 3.

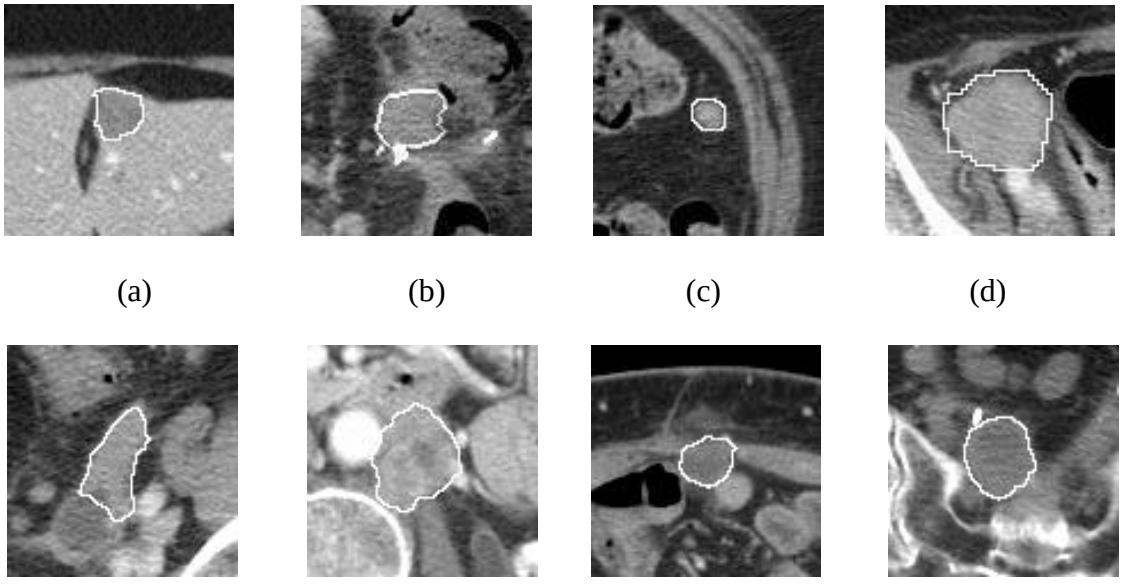


Figure 3: Segmentation samples of (a) connected (b) neighborhood (c) canny (d) multi-canny (e) fast marching (f) shape detection (g) geometric (h) threshold segmentation algorithms

Shape/volume based features

This group includes 10 shape and/or volume based features (F1-F10) [14-17]. These features estimate tumor characteristics including shape, volume, and related geometric distortions.

1. **Convexity:** It describes the smoothness of edges of the tumor, which can be calculated as follows:

$$convexity = \frac{Tumor\ region\ area}{Convex\ region\ area} \tag{14}$$

2. **Tumor volume:** This is the total volume of all the tumor voxels.
3. **Max radius:** we computed all the possible radii between the center and all tumor surface pixels. Then the maximum of radii is considered as the max radius.

4. **STD RL:** It is the ratio of standard deviation and average of all the radii between center and tumor surface pixels.

$$STD\ RL = \frac{\sqrt{\frac{\sum_{i=1}^{N}(l_i - \overline{R}_i)^2}{N-1}}}{\frac{1}{N}\sum_{i=1}^{N} R_i} = \frac{STD\ of\ radii}{mean\ of\ radii} \tag{15}$$

5. **Surface area:** In order to simplify the computation, the tumor is approximated by generalized cylinder. Accordingly, the surface area of the generalized cylinder is used as an estimation of the tumor surface area.

6. **Compactness1:**

$$compactness1 = \frac{V}{\sqrt{\pi} A^{\frac{2}{3}}} \tag{16}$$

7. **Compactness2:**

$$compactness2 = \frac{36\,\pi V^2}{A^3} \tag{17}$$

8. **Spherical disproportion:** It measures the difference between the tumor surface and a sphere with equivalent radius:

$$spherical\ disproportion = \frac{A}{4\pi R^2} \tag{18}$$

9. **Sphericity:** This feature is a measurement of the shape distortion of the tumor when comparing with the sphere, which is computed as follows:

$$Sphericity = \frac{\pi^{\frac{1}{3}}(6V)^{\frac{2}{3}}}{A} \tag{19}$$

10. **Surface area to tumor volume:** It is the ratio of surface area to the tumor volume.

Density features

These features are computed based on pixel intensities within the tumor and/or surrounding boundary pixels and describe their distribution [18]. This subgroup consists of 21 features (F11-F31), among which (F11-F15) are computed based on the 2D central slice and the rest of them are computed based on 3D tumors which are discussed below:

1. **ISO-intensity:** The feature computes the density of all the outer ring (boundary) pixels for which the intensity is higher than the threshold determined by the pixels inside the tumor. (where,10^{th} percentile intensity of pixel values inside the tumor is the threshold)
2. **Fluctuation mean:** The fluctuation is defined as the maximum absolute difference between the target (Central) pixel and all the neighbor pixels within its 5×5 neighborhood. Thus fluctuation mean is the average fluctuation value for all the tumor pixels.
3. **Fluctuation standard deviation:** This feature is the standard deviation of fluctuation value of all the tumor pixels.
4. **Gradient Mean:** It is defined as the average gradient value of all the tumor boundary pixels.
5. **Gradient standard deviation:** This feature is the standard deviation of gradient value of all boundary pixels.
6. **Density:** This feature computes the average intensity of all the pixel values within the tumor.
7. **Standard deviation of density:** It computes the standard deviation of all the pixel values within the tumor.

8. **Mean contrast:** For this feature, we initially compute an inner ring within the tumor and an outer ring surrounding the tumor then, the intensity ratio of the inner ring to surrounding outer ring is computed as follows:

$$mean\ contrast = \frac{mean\ of\ inner\ ring\ tumor\ pixels\ (\overline{I_{inner}})}{mean\ of\ outer\ ring\ boundary\ pixels\ (\overline{I_{outer}})} \quad (20)$$

9. **Contrast:** It is the difference between the mean of inner ring tumor pixels and the mean of surrounding outer ring boundary pixels.

$$contrast = \overline{I_{inner}} - \overline{I_{outer}} \quad (21)$$

10. **Skewness:** Skewness measures the asymmetry of the probability distribution curve of all the tumor pixel values about its mean.

$$skewness = \sqrt{N} * \frac{\sum_{i=1}^{N}(I_i - \bar{I})^3}{\sum_{i=1}^{N}(I_i - \bar{I})^{\frac{3}{2}}} \quad (22)$$

11. **Kurtosis:** Kurtosis measures the "tailed-ness" of the tumor density distribution when comparing to the standard normal distribution:

$$kurtosis = N * \frac{\sum_{i=1}^{N}(I_i - \bar{I})^4}{(\sum_{i=1}^{N}(I_i - \bar{I})^2)^2} \quad (23)$$

12. **STD ratio:** It is defined as the ratio of STD of tumor intensity to the boundary intensity.

13. **Energy:** It is a sum of the squared tumor pixel values.

$$Energy = \sum_{i=1}^{N} I_i^2 \quad (24)$$

14. **Entropy:** This feature describes the randomness/ uncertainty in an image.

$$Entropy = \sum_{i=1}^{N_l} P_i \log_2 P_i \quad (25)$$

Where P is first order histogram of tumor pixels with N_l discrete intensity levels.

15. **Max:** It is the maximum value of all the tumor pixels
16. **Mean absolute deviation:** It is defined as the mean absolute deviation between the tumor pixel value and the average tumor intensity:

$$Mean\ absolute\ deviation = \frac{1}{N}\sum_{i=1}^{N} |I_i - \bar{I}| \quad (26)$$

17. **Median:** It is the median value of all the tumor pixels.
18. **Min:** It is the minimum value of all the tumor pixels.
19. **Range:** It measures difference between the maximum and a minimum value of all tumor pixels.
20. **RMS:** it is the root mean square value of tumor pixels.
21. **Uniformity:** This feature is a measure of histogram randomness and can be computed as follows:

$$Uniformity = \sum_{i=1}^{N_l} P_i^2 \quad (27)$$

Texture based features

These features are computed using the gray level run length (GLRL). Gray level run is the number of consecutive elements with the same gray level in a certain direction (i.e. 0°, 45°, 90°, and 135°), while the run length is the number of the gray level run.

The run length provides meaningful texture information for tumor classification [19-20]. In our study, we estimated a total of 11 texture features [21-23] and are as follows:

1. **Short Run Emphasis (SRE)**

$$SRE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)}{j^2} = \frac{1}{n_r}\sum_{j=1}^{N}\frac{p_r(j)}{j^2} \quad (28)$$

2. **Long Run Emphasis (LRE)**

$$LRE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}p(i,j).j^2 = \frac{1}{n_r}\sum_{j=1}^{N}p_r(j).j^2 \quad (29)$$

3. **Gray-Level Non-uniformity (GLN)**

$$GLN = \frac{1}{n_r}\sum_{i=1}^{M}\left(\sum_{j=1}^{N}p(i,j)\right)^2 = \frac{1}{n_r}\sum_{i=1}^{M}p_g(i)^2 \quad (30)$$

4. **Run-Length Non-uniformity (RLN)**

$$RLN = \frac{1}{n_r}\sum_{j=1}^{N}(\sum_{i=1}^{M}p(i,j))^2 = \frac{1}{n_r}\sum_{j=1}^{N}p_r(i)^2 \quad (31)$$

5. **Run Percentage (RP)**

$$RP = \frac{n_r}{n_p} \quad (32)$$

6. **Low Gray-Level Run Emphasis (LGRE)**

$$LGRE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)}{i^2} = \frac{1}{n_r}\sum_{i=1}^{M}\frac{p_g(i)}{i^2} \quad (33)$$

7. **High Gray-Level Run Emphasis (HGRE)**

$$HGRE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}p(i,j).i^2 = \frac{1}{n_r}\sum_{i=1}^{M}p_g(i).i^2 \quad (34)$$

8. **Short Run Low Gray-Level Emphasis (SRLGE)**

$$SRLGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)}{i^2.j^2} \quad (35)$$

9. **Short Run High Gray-Level Emphasis (SRHGE)**

$$SRHGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j).i^2}{j^2} \quad (36)$$

10. **Long Run Low Gray-Level Emphasis (LRLGE)**

$$LRLGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j).j^2}{i^2} \quad (37)$$

11. Long Run High Gray-Level Emphasis (LRHGE)

$$LRHGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} p(i,j).i^2.j^2 \quad (38)$$

In these formulas, P is a GLRL matrix, in which P (i, j) is the number of runs for gray level i and length j. n_r is total number of runs, and n_p is the number of distinctive elements in the GLRL matrix P.

Given that the above 11 GLRL features can be estimated in four different directions (i.e. 0°, 45°, 90° and 135°), thus we finally have 44 GLRL features. (F32-F75 GLRL features in four directions).

Wavelet features

In this study, a two-dimensional wavelet transform ("Coiflet 1" filter) was applied on each image [18], which decomposes the original image (I) into four components: I_{LL}, I_{LH}, I_{HL}, and I_{HH}, where L and H denotes the low and high pass filters respectively. For example, I_{HL} is obtained by applying high-pass filter along the x-direction and low pass filter along y-direction on original CT image (I). Mathematically, it can be described as follows:

$$I_{HL}(i,j) = \sum_{p=1}^{N_H}\sum_{q=1}^{N_L} H(p)L(q)I(i+p,j+q) \quad (39)$$

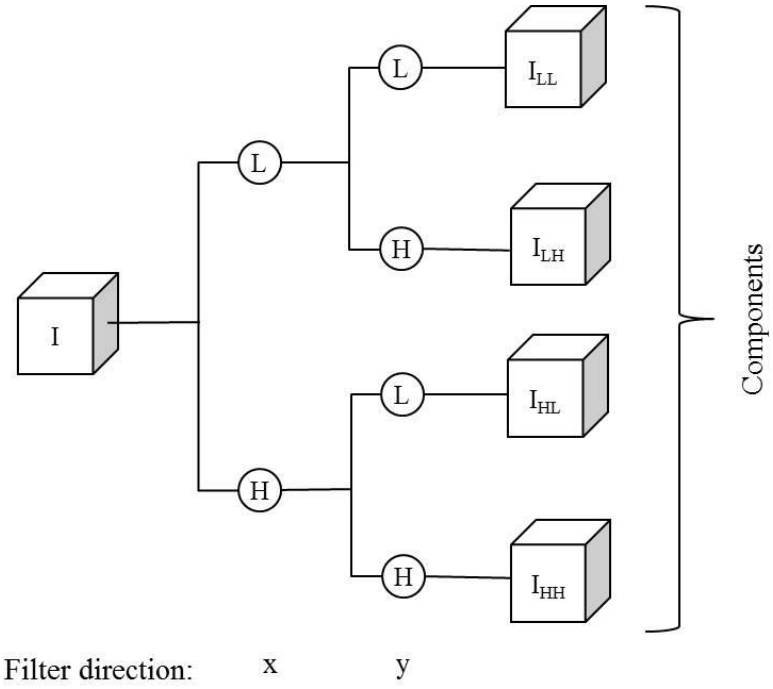


Figure 4: Wavelet decomposition of target CT slice using wavelet transforms (one-level and un-decimated two-dimensional wavelet transforms using "Coiflet 1" filter)

Accordingly, for each component, we estimated the same density features as described above. All the features are summarized in Table 2.

Table 2: Summary of all the 159 features separated by their feature class

Feature class	Feature number	Feature description
Shape	1-10	convexity, Tumor volume, maximum radius, STD RL , surface area, compactness feature 1, compactness feature 2, spherical disproportion, Sphericity, ratio of surface area to Tumor volume
Density	11-31	ISO-intensity, fluctuation mean, fluctuation STD, gradient mean, gradient STD, Density, density STD, mean contrast, contrast, skewness, kurtosis, STD ratio of tumor to boundary, energy, entropy, maximum intensity, mean absolute deviation, median, minimum, range, rms, uniformity
Texture	32-75	11 gray level run length based features extracted in four directions (0°, 45°, 90°, and 135°)
Wavelet	76-159	Apply the density features on the four wavelet components

For each case, we computed these 159 features on both pre-treatment (F_i^{Pre}) and 4-6 week follow-up (F_i^{Post}) CT image sets. Then, the tumor feature differences $\Delta F_i = F_i^{Post} - F_i^{Pre}$ were also computed. After that, feature sets $F_i^{Pre}, \Delta F_i$ were normalized within the range between $\mu + 2\sigma$ and $\mu - 2\sigma$, where μ and σ denote the mean and standard deviation of the feature values respectively. In accordance with the RECIST guidelines, 1 to 5 tumors were tracked and studied by radiologists for different cases. In this study, the average feature computed from M tumors ($1 \leq M \leq 5$) corresponding to the same case was also computed and used to represent the final case-based feature value. We used the case-based features to assess their association to the 6-month PFS of the patient.

After building two initial feature pools containing F_i^{Pre} and $\Delta F_i = F_i^{Post} - F_i^{Pre}$, we firstly assessed the performance of each feature, using receiver operating characteristic (ROC) curve[21-22]. Then we sorted features based the area under the curve (AUC) values and top 20 features are selected in both feature pools, to investigate whether using feature difference $\Delta F_i = F_i^{Post} - F_i^{Pre}$ could yield significantly high performance than using the features computed from pre-therapy CT images only. Next, an equal-weighted Fusion model was built to combine the optimal feature cluster and generate a new combined feature marker to predict the 6-month progression free survival (PFS). We also used the ROC analysis method to assess the predicting performance of the combined marker. Finally, we built three confusion matrices using the prediction results generated by two fusion image markers and RECIST criteria used by radiologists in current clinical practice. The predicting accuracy of our new quantitative image markers were compared with the conventional RECIST method.

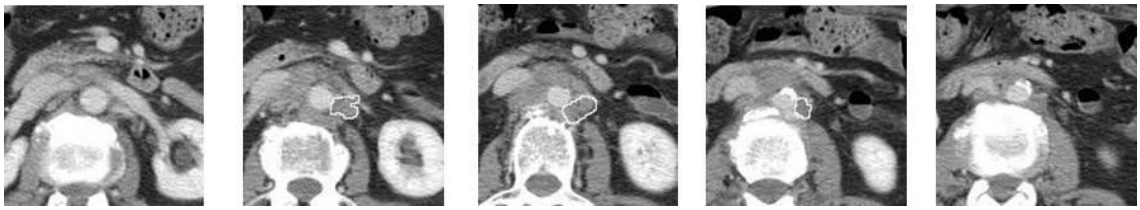
Chapter 4: Results

Tumor segmentation results

A number of 138 tumors are used to assess the performance of these segmenting algorithms. All of these tumors are selected and tracked from the advanced ovarian cancer patients with recurrent metastatic tumors. Figure 3 illustrates some typical segmenting examples of the eight methods, which demonstrated that these methods are effective for the tumor segmentation of the ovarian cancer cases. However, given that the metastatic tumors are highly sophisticated, none of the above eight approaches are able to effectively segment all the tumors. As indicated in Table 3, the PDE based methods are achieved a better performance than the region growing and Canny boundary operator methods, for which the fast marching, geometric active contour, and threshold level set yielded a number of 106, 104, and 107 satisfied segmenting results. In addition, although each single method is not able to segment all of these tumors, each of the tested tumors can be successfully segmented by at least one of the eight methods without the manual boundary correction.

Table 3: Summary of the eight segmenting result of the 138 metastatic tumors for the advanced stage ovarian cancer patients

Segmenting methods	Region growing	Neighborhood region growing	Canny operator	Multi-scale canny
Unsatisfied	78	102	76	109
Satisfied	60	36	62	29
Segmenting methods	Fast marching	Geometric active contour	Shape detection	Threshold level set
Unsatisfied	32	34	48	31
Satisfied	106	104	90	107



(a) (b) (c) (d) (e)

Figure 5: An aortic tumor segmented by our hybrid segmenting scheme from the (a) 2nd upper adjacent slice (b) 1st upper adjacent slice (c) central slice selected by the radiologist (d) 1st lower adjacent slice (e) 2nd lower adjacent slice

In order to segment the 3D tumor from the multiple CT slices, the professional radiologist will first mark one central slice containing the tumor, which was segmented by one of the developed algorithms. Once the tumor region and boundary contour are segmented on this center slice, the result provides a reference to guide the segmentation of the tumor regions depicting on the two adjacent CT image slices. The scheme will keep segmenting the tumors on the adjacent slices until the tumor disappears. Figure 5 shows an example of 3-D aortic tumor segmented from a series of adjacent CT image slices. The tumor depicted on Figure 5 (c) was first segmented from the target slice marked by the radiologist, which has the largest 2-D area. Figure 5 (d) and (e) indicate that the tumor significantly decreases and disappear on the first and second lower adjacent slices, respectively. The similar results can also be revealed on the upper adjacent slices, as shown in Figure 5 (a) and (b).

Quantitative image feature analysis results

Table4, Table 5 summarizes two sets of 20 best performed image features computed from pre-therapy CT image only and two CT scans acquired pre- and pre-post therapy. The

pre-therapy set consists of five tumor density features (Skewness, Uniformity, Entropy, kurtosis, and STD), thirteen wavelet features (Contrast-LH, Contrast-HL, Skewness-HH, Density std Ratio-HL, Kurtosis-HL, Mean contrast-LH, Density std Ratio-HH, Skewness-LH, Range-HL, Density-LH, Mean contrast-HL, Minimum density-LH, and Kurtosis-LL), and two tumor shape features (Sphericity and Compactness-2). AUC values of these 20 features range from 0.5692 to 0.6842. The best performed feature is Skewness with an AUC value of 0.6842±0.0557. Similarly, another set based on the image feature difference between the pre- and post-therapy has nine tumor density features (Energy, Median, Rms, Average density, Maximum density, Density range, Fluctuation std, std-RL, and Entropy), eight wavelet features (Fluctuation std-LL, Entropy-HL, Energy-HH, Entropy-LL, Energy-HL, Maximum density-HH, Maximum density-HL, and Gradient mean-HL), and three tumor shape features (Compactness-1, Tumor volume, and Surface area). AUC values of using these 20 features ranged from 0.6511 to 0.7705, among which the feature of Compactness-1 yielded the highest AUC value of 0.7705±0.0495.

Table 4: Summary of top 20 features obtained using pretreatment CT images

Feature	AUC ± STD	Feature	AUC ± STD
Skewness	0.6842 ± 0.0557	Mean contrast-LH	0.5960 ± 0.0601
Contrast-LH	0.6521 ± 0.0592	Kurtosis	0.5923 ± 0.060
Contrast-HL	0.6426 ± 0.0576	Density std Ratio-HH	0.5907 ± 0.0614
Uniformity	0.6367 ± 0.0579	Skewness-LH	0.5877 ± 0.0597
Skewness-HH	0.6225 ± 0.0586	Range-HL	0.583 ± 0.0611
Sphericity	0.6209 ± 0.0579	Density-LH	0.578 ± 0.0609
Entropy	0.6192 ± 0.0589	Mean contrast-HL	0.5774 ± 0.0599
Density std Ratio-HL	0.6178 ± 0.0581	Minimum density-LH	0.5734 ± 0.0617
Kurtosis-HL	0.6149 ± 0.0587	Kurtosis-LL	0.5704 ± 0.0608
Compactness2	0.6022 ± 0.0585	STD of density	0.5692 ± 0.0613

Table 5: Summary of top 20 features obtained using both pre and posttreatment CT images

Feature	AUC ± STD	Feature	AUC ± STD
Compactness 1	0.7705 ± 0.0495	Density range	0.677 ± 0.0558
Volume	0.7547 ± 0.0512	Energy-HH	0.6696 ± 0.0567
Surface area	0.7389 ± 0.0523	Entropy-LL	0.6688 ± 0.0555
Fluctuation std-LL	0.7108 ± 0.0535	Energy-HL	0.6671 ± 0.0597
Energy	0.7066 ± 0.0548	Maximum density-HH	0.6629 ± 0.058
Median	0.6949 ± 0.0542	Fluctuation std	0.6616 ± 0.056
Entropy-HL	0.6865 ± 0.0545	Maximum density-HL	0.6603 ± 0.0582
Rms	0.685 ± 0.0549	Gradient mean-HL	0.6583 ± 0.0565
Density	0.6835 ± 0.0549	Std-RL	0.6532 ± 0.0563
Maximum density	0.6815 ± 0.0555	Entropy	0.6511 ± 0.0566

The selected optimal clusters consisting of 4 and 9 features from two feature sets of (1) using pretreatment CT images only (F_i^{Pre}) and (2) using both pre- and post-treatment CT images (ΔF_i) respectively. These optimal clusters were evaluated using ROC curve. As indicated in Figure 6Figure 7, the AUC values of final fusion markers are 0.8103 ± 0.0447 and 0.8292 ± 0.0431 for optimal feature cluster sets (F_i^{Pre}) and(ΔF_i) respectively. The predicting powers of the fused markers are significantly higher when comparing to the AUC values of single features in the initial individual feature sets. At a specificity of 0.6, the sensitivity values are 0.8241 and 0.8735 for the final pre-treatment and the pre-post treatment fusion markers respectively.

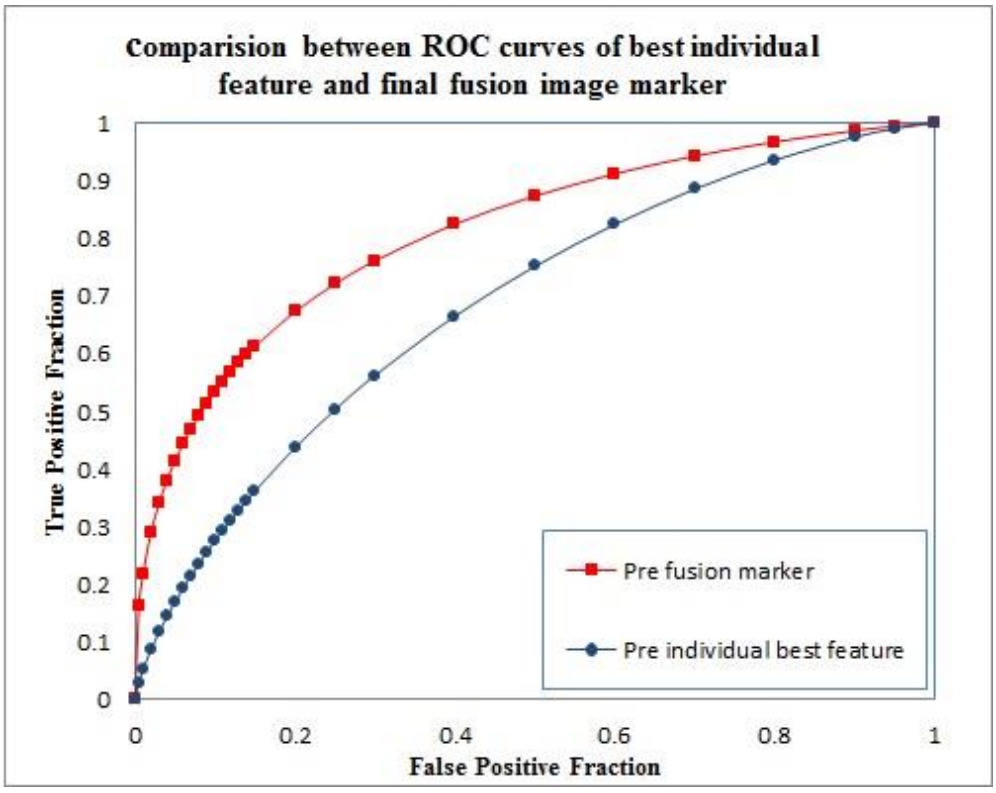


Figure 6: ROC curves of the best individual feature and the final fusion image marker for pre-treatment features

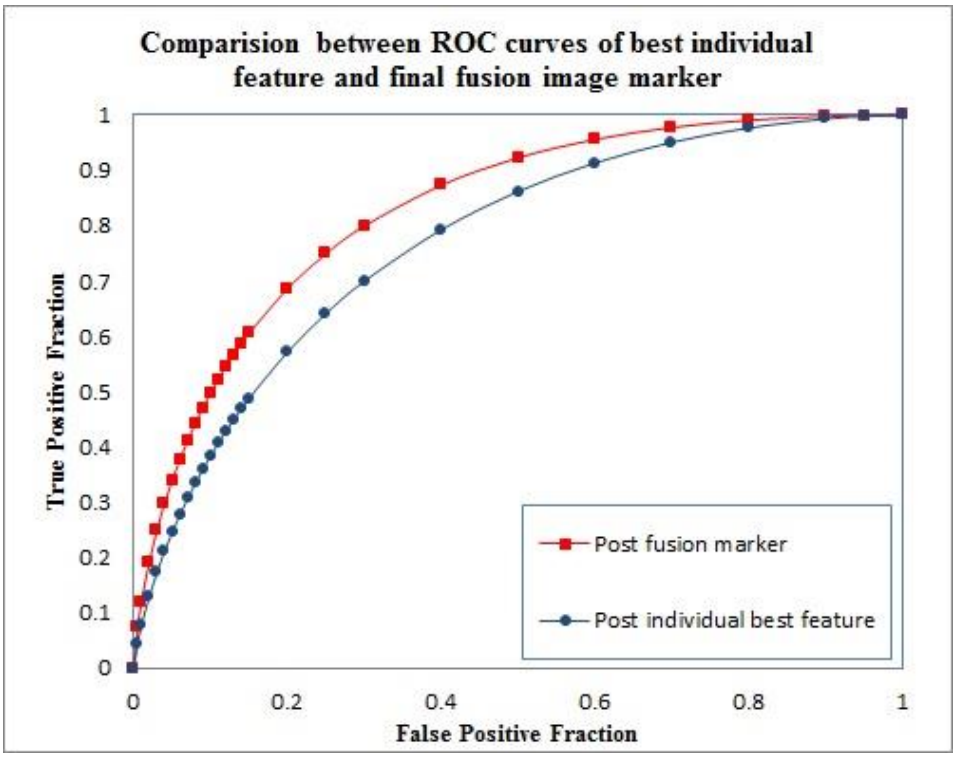


Figure 7: ROC curves of the best individual feature and the final fusion image marker for pre-posttreatment features

Table 6,Table 7, and Table 8 shows the confusion matrices obtained using two fusion markers and conventional RECIST criteria respectively. The pre-treatment marker predicts 46 cases as responsive cases, among which 36 cases are clinically responsive to the therapy. Similarly, this marker predicts 45 cases as non-responsive cases, and 29 of them are clinical non-responders. Thus the pre-treatment marker yields a positive predictive value (PPV) of 78.26% and a negative predictive value (NPV) of 64.44%. The overall prediction accuracy of this pre-fusion marker is achieved as 71.43%. For the pre-post treatment marker, 60 and 31 cases are predicted as responsive and non-responsive cases, among which 47 and 26 cases are clinical responders or non-responders. Therefore the PPV and NPV are 78.33% and 83.87%, respectively. As a result, an overall prediction

accuracy of 80.22% was obtained for the pre-post fusion marker. As a comparison, Table 8 summarizes the confusion matrix of the conventional RECIST method, in which CR, PR, and SD are considered as responsive to treatment by maintaining 6-month PFS and PD as non-responsive to treatment. The overall prediction accuracy using RECIST criteria is 74%

Table 6: A confusion matrix predicting 6-month PFS when fusing features obtained from pre-treatment CT images

6-month PFS / Prediction	Yes	No
Yes	36	10
No	16	29

Table 7: A confusion matrix predicting 6-month PFS when fusing features obtained from both pre- and post-treatment CT images

6-month PFS / Prediction	Yes	No
Yes	47	13
No	5	26

Table 8: A confusion matrix predicting 6-month PFS when using RECIST criteria

6-month PFS / Prediction	Yes	No
Yes	52	23
No	0	16

Chapter 5: Discussion

CT is the most widely used imaging modality for assessing the efficacy of chemotherapy drugs in treating ovarian cancer patients. Currently, most of the clinical evaluations are based on RECIST criteria which use one-dimensional tumor diameter change to categorize patients' response to chemotherapy. However, the previous studies have proved that diameter change cannot be sole measure for optimal assessment. For example, 1) when using bevacizumab-based therapy in clinical trials to treat recurrent ovarian cancer, RECIST based assessment categorized 21% as responsive when there were actually 40% patients who remained 6-moth PFS [26]. 2) When using immune-stimulatory agents like CTLA4 for therapy, some tumors might have an initial increase in tumor size on CT images which is not permanent and can shrink at a later time [27-28]. As a result, there is a need for new approaches which is not entirely based on one-dimensional size.

In this study, we developed a CAD model which computes a number of 159 quantitative image features to estimate the tumor shape, size, density and texture characteristics. Two sets of the quantitative features, pre-treatment and pre-post treatment features, were computed by our new CAD scheme to predict the response of ovarian cancer patients to chemotherapy. The results indicate that the pre-and pre-post treatment features are able to yield AUC values of 0.8103 ± 0.0447 and 0.8292 ± 0.0431 respectively, which is significantly higher than the results accomplished by the conventional RECIST method. The performance superiority can be attributed by two main advantages of our new method.

First, this study indicates that pre-therapy CT imaging data includes useful information for the therapy response prediction. The study supports the concept of emerging Radiomics, which hypothesizes that the tumor heterogeneity features detected and quantified from radiographic images (e.g., CT) enable to phenotype corresponding genomic or biological biomarkers to predict cancer prognosis or tumor response to the therapies [18]. In this study, we found that a number of tumor density and texture heterogeneity related image features, had significantly higher discriminatory or prediction power than random guess (AUC = 0.5). As indicated in Table 4, the highest AUC value of 0.6842 ± 0.0557 was achieved using one image feature computed from the pre-therapy CT images. Furthermore, by fusing a cluster of four selected optimal features using a simple equal-weighting method, the prediction performance significantly increased to 0.8103±0.0447 ($p < 0.05$). After applying an operation threshold, this fusion image marker achieved an overall prediction accuracy of 71%. In addition, this is a totally different approach of using RECIST in current clinical practice, which requires comparison of tumor size change in two sets of CT images acquired pre- and post-therapy. Thus, one potential advantage of using quantitative image feature analysis is enabling to predict or assess cancer prognosis or PFS before therapy. If successful, this will help clinicians (e.g., oncologists) select optimal or personized cancer treatment strategy for the individual patients to achieve the maximum therapy benefit while minimizing unnecessary toxicity.

Second, our study also showed that although using pre-therapy CT images can help predict patients' response to therapy, adding quantitative images features computed from post-therapy CT images could provide more discriminatory information. As a result,

many individual image features computed from pre- and post-therapy CT images yielded significantly higher prediction performance than using the image features computed from pre-therapy CT images only. In addition, comparing to using RECIST which yielded 74% prediction accuracy in this study, the overall prediction accuracy of using quantitative imaging marker increased to 80%. The higher performance is probably due to the fact that quantitative image feature analysis method is able to extract more useful information of tumor characteristic change from the CT images (i.e., tumor volume, density, boundary spiculation). RECIST, however, only assesses one dimension tumor size change from two sets of CT images. As shown in Table 5, besides tumor volume change, other 19 selected features with higher discriminatory power cover more tumor characteristics related to tumor shape (compactness-1) and density heterogeneity changes. This can be attributed by the fact that the chemotherapy treatment will not only lead to the change of the tumor size, but also result in the change of some other tumor characteristics, such as the tumor density, necrosis, stiffness and other heterogeneity patterns. Many of these characteristic changes will not be reflected on the tumor size measurement, but can be extracted by quantitative image feature analysis. Because of these facts, quantitative image feature analysis is a new promising approach to more accurately predict patients' response to therapies.

Even though the results are encouraging, this study has some limitations. First, we noticed that although adding the post-therapy CT image data can significantly improve the predicting performance on the single feature (e.g., 0.7705 ± 0.0495 vs 0.6842 ± 0.0557), the performance improvement between using two fusion based image markers was not substantial (e.g., 0.8292 ± 0.0431 vs 0.8103 ± 0.0447). The reason behind this

observation needs to be further investigated using other independent image datasets. Second, due to the limited size of our image dataset, only a simple equal weighted fusion method was applied to build image marker. In the future study, a large image dataset needs to be built and the state of the art machine learning methods need to be investigated to optimally combine image features to further improve the predicting performance. Third, for each case, we only investigated the tumors marked and tracked by the radiologist based on RECIST guideline. For the features of these different tracked tumors, we only computed the average value as the final case-based feature, which might not be the optimal method. Thus, we still need to investigate the more effective algorithms to quantify the case-based image features. Fourth, in this dataset, the patient cases were not classified with different chemotherapy drugs. Given that the tumor responsive to a certain type of drug may contain some specific tumor characteristics, dividing and finding the best features for the patients with different therapy groups may further improve the predicting accuracy. Last, although we have accumulated 91 cancer cases, the robustness of our experiment results need to be tested and verified using large and diverse image datasets in the future.

In conclusion, even with few limitations, we believe that this initial study is unique and valid, which demonstrated that pre-therapy CT image analysis provide useful information in predicting ovarian cancer patients response to chemotherapy and addition of post-therapy information can further improve the model effectiveness by generating higher prediction accuracy than both pre-therapy and conventional RECIST criteria methods. As a result, this study provides a valid foundation for us to continue developing more robust quantitative image analysis scheme to identify clinically useful imaging

markers, which can more accurately predict the clinical benefit of the chemotherapy at the early stage for ovarian and/or other cancer patients.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA Cancer J Clin*, vol. 66, no. 1, pp. 7–30, 2016.

[2] S. Sato and H. Itamochi, "Neoadjuvant chemotherapy in advanced ovarian cancer: latest results and place in therapy.," *Ther. Adv. Med. Oncol.*, vol. 6, no. 6, pp. 293–304, 2014.

[3] L. A. Baldwin *et al.*, "Ten-Year Relative Survival for Epithelial Ovarian Cancer," in *Obstet Gynecol*, vol. 120, 2012, pp. 612–618.

[4] D. Bell *et al.*, "Integrated genomic analyses of ovarian carcinoma," *Nature*, vol. 474, no. 7353, pp. 609–615, 2011.

[5] C. M. Tempany, K. H. Zou, S. G. Silverman, D. L. Brown, a B. Kurtz, and B. J. McNeil, "Staging of advanced ovarian cancer: comparison of imaging modalities--report from the Radiological Diagnostic Oncology Group.," *Radiology*, vol. 215, no. 3, pp. 761–767, 2000.

[6] P. Gu, L.-L. Pan, S.-Q. Wu, L. Sun, and G. Huang, "CA 125, PET alone, PET-CT, CT and MRI in diagnosing recurrent ovarian carcinoma: a systematic review and meta-analysis.," *Eur. J. Radiol.*, vol. 71, no. 1, pp. 164–74, 2009.

[7] S. Kyriazi, S. B. Kaye, and N. M. DeSouza, "Imaging ovarian cancer and peritoneal metastases--current and emerging techniques.," *Nat. Rev. Clin. Oncol.*, vol. 7, no. 7, pp. 381–93, 2010.

[8] P. Therasse *et al.*, "New Guidelines to Evaluate the Response to Treatment in Solid Tumors," *J. Natl. Cancer Inst.*, vol. 92, no. 3, pp. 205–216, 2000.

[9] E. A. Eisenhauer *et al.*, "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)," *Eur. J. Cancer*, vol. 45, no. 2, pp. 228–247, 2009.

[10] M. R. Sharma, M. L. Maitland, M. J. Ratain, A. T. Fojo, and A. Noonan, "RECIST: No longer the sharpest tool in the oncology clinical trials toolbox -

Point/Counterpoint," *Cancer Research*, vol. 72, no. 20. pp. 5145–5150, 2012.

[11] R. G. Abramson, C. R. McGhee, N. Lakomkin, and C. L. Arteaga, "Pitfalls in RECIST Data Extraction for Clinical Trials: Beyond the Basics," *Academic Radiology*, vol. 22, no. 6. pp. 779–786, 2015.

[12] L. J. Fallowfield, L. J. Fallowfield, and A. Fleissig, "The value of progression-free survival to patients with advanced-stage cancer," *Nat. Publ. Gr.*, vol. 9, no. 9, pp. 41–4710, 2011.

[13] Y. Qiu *et al.*, "Early prediction of clinical benefit of treating ovarian cancer using quantitative CT image feature analysis.," *Acta Radiol.*, 2015.

[14] B. Zheng, Y. H. Chang, and D. Gur, "On the reporting of mass contrast in CAD research.," *Med. Phys.*, vol. 23, no. 12, pp. 2007–9, 1996.

[15] G. M. te Brake, N. Karssemeijer, and J. H. C. L. Hendriks, "An automatic method to discriminate malignant masses from normal tissue in digital mammograms," *Phys. Med. Biol.*, vol. 45, no. 10, pp. 2843–2857, 2000.

[16] G. M. Brake, N. Karssemeijer, and J. H. C. L. Hendriks, "An automatic method to discriminate malignant masses from normal tissue in digital mammograms An automatic method to discriminate malignant masses from normal tissue in digital mammograms," *Phys. Med. Biol.*, vol. 45, pp. 2843–2587, 2000.

[17] B. Zheng *et al.*, "A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment.," *Med. Phys.*, vol. 33, no. 1, pp. 111–117, 2006.

[18] H. J. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat Commun*, vol. 5, p. 4006, 2014.

[19] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *Syst. Man Cybern. IEEE Trans.*, no. 6, pp. 610–621, 1973.

[20] X. Tang, "Texture information in run-length matrices," *IEEE Trans. Image*

Process., vol. 7, no. 11, pp. 1602–1609, 1998.

[21] M. M. Galloway, "Texture analysis using gray level run lengths," *Comput. Graph. Image Process.*, vol. 4, no. 2, pp. 172–179, 1975.

[22] A. Chu, C. M. Sehgal, and J. F. Greenleaf, "Use of gray value distribution of run lengths for texture analysis," *Pattern Recognit. Lett.*, vol. 11, no. 6, pp. 415–419, 1990.

[23] B. V. Dasarathy and E. B. Holder, "Image characterizations based on joint gray level-run length distributions," *Pattern Recognit. Lett.*, vol. 12, no. 8, pp. 497–502, 1991.

[24] B. Zheng, "Computer-Aided Detection in Mammography : An Assessment of Performance on Current and Prior Images," *Acad. Radiol.*, vol. 9, no. 11, pp. 1245–1250, 2002.

[25] C. E. Metz, B. A. Herman, and J.-H. Shen, "Maximum Likelihood Estimation of Receiver Operating Characteristic (Roc) Curves From Continuously-Distributed Data," *Ltd. Revis. May Stat. Med. Stat. Med*, vol. 17, no. October 1996, pp. 1033–1053, 1996.

[26] R. a Burger, M. W. Sill, B. J. Monk, B. E. Greer, and J. I. Sorosky, "Phase II trial of bevacizumab in persistent or recurrent epithelial ovarian cancer or primary peritoneal cancer: a Gynecologic Oncology Group Study," *J. Clin. Oncol.*, vol. 25, no. 33, pp. 5165–71, 2007.

[27] J. D. Wolchok *et al.*, "Guidelines for the evaluation of immune therapy activity in solid tumors: Immune-related response criteria," *Clin. Cancer Res.*, vol. 15, no. 23, pp. 7412–7420, 2009.

[28] A. Hoos *et al.*, "Improved endpoints for cancer immunotherapy trials," *J. Natl. Cancer Inst.*, vol. 102, no. 18, pp. 1388–1397, 2010.