# Categorizing Extent of Tumor Cell Death Response to Cancer Therapy Using Quantitative Ultrasound Spectroscopy and Maximum Mean Discrepancy

Mehrdad J. Gangeh, *Member, IEEE*, Ali Sadeghi-Naini, *Member, IEEE*, Michael Diu, *Student Member, IEEE*, Hadi Tadayyon, *Student Member, IEEE*, Mohamed S. Kamel, *Fellow, IEEE*, and Gregory J. Czarnota

*Abstract*—Quantitative ultrasound (QUS) spectroscopic techniques in conjunction with maximum mean discrepancy (MMD) have been proposed to detect, and to classify noninvasively the levels of cell death in response to cancer therapy administration in tumor models. Evaluation of xenograft tumor responses to cancer treatments were carried out using conventional-frequency ultrasound at different times after chemotherapy exposure. Ultrasound data were analyzed using spectroscopic techniques and multi-parametric QUS spectral maps were generated. MMD was applied as a distance criterion, measuring alterations in each tumor in response to chemotherapy, and the extent of cell death was classified into less/more than 20% and 40% categories. Statistically significant differences were observed between "pre-" and "post-treatment" groups at different times after chemotherapy exposure, suggesting a high capability of proposed framework for detecting tumor response noninvasively. Promising results were also obtained for categorizing the extent of cell death response in each tumor using the proposed framework, with gold standard histological quantification of cell death as ground truth. The best classification results were obtained using MMD when applied on histograms of QUS parametric maps. In this case, classification accuracies of 84.7% and 88.2% were achieved for categorizing extent of tumor cell death into less/more than 20% and 40%, respectively.

*Index Terms*—Cancer treatment, classification methods, kernel methods, nonparametric methods, personalized medicine, quantitative ultrasound.

*M. J. Gangeh is with the Departments of Medical Biophysics, and Radiation Oncology, University of Toronto, Toronto, ON, M5G 2M9 Canada, also with the Departments of Radiation Oncology, and Imaging Research—Physical Sciences, Sunnybrook Health Sciences Centre, Toronto, ON, M4N 3M5 Canada (e-mail: mehrdad.gangeh@utoronto.ca).

A. Sadeghi-Naini and G. J. Czarnota are with the Departments of Medical Biophysics, and Radiation Oncology, University of Toronto, Toronto, ON, M5G 2M9 Canada, also with the Departments of Radiation Oncology, and Imaging Research—Physical Sciences, Sunnybrook Health Sciences Centre, Toronto, ON, M4N 3M5 Canada (e-mail: a.sadeghi.naini@utoronto.ca; gregory.czarnota@sunnybrook.ca).

M. Diu and M. S. Kamel are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1 Canada (e-mail: mdiu@uwaterloo.ca; mkamel@pami.uwaterloo.ca).

H. Tadayyon is with the Department of Medical Biophysics, University of Toronto, Toronto, ON, M5G 2M9 Canada, and also with the Department of Physical Sciences, Sunnybrook Health Sciences Centre, Toronto, ON, M4N 3M5 Canada (e-mail: hadi.tadayyon@sunnybrook.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2014.2312254

## I. INTRODUCTION

PERSONALIZED cancer therapy has recently attracted much attention among researchers in various areas of science and engineering including biomedical engineering, biophysics, medicine, and biology [1], [2], [3]. Motivated by high rates of poor responses of patients to standard routine treatments, personalized cancer therapy aims at altering therapies for refractory patients on an individual basis. Utilizing faster feedback mechanisms, personalized cancer therapy is expected to avoid needless and/or harmful continuation of an ineffective cancer treatment, or even facilitate switching to a salvage therapy early on during the course of treatment.

The monitoring of cancer therapy effects is broadly performed at two different levels: macroscopic/anatomical or microscopic/cellular. The former mainly relies on the classical clinical practice to measure the size of tumor, which has been established for research, clinical screening, and treatment planning; see [4] for a comprehensive survey. The latter, i.e., monitoring tumor response to treatment at microscopic/cellular level [5], [6], [7], is a much newer field. The impetus for research in this area is to shift away from the conventional paradigm of using tumor size changes as a measure of treatment effectiveness. These changes in size can take weeks to months to become apparent, and do not always occur even when the treatment is effective [5], [8].

Functional imaging modalities, such as single photon emission computed tomography (SPECT), positron emission tomography (PET), and magnetic resonance imaging (MRI) play an important role as they can potentially enable noninvasive evaluation of response to cancer therapies at microscopic level early after the start of treatment (days as opposed to months in standard clinical practice) [5], [8], [7]. However, these methods have the disadvantages of being expensive and requiring injection of exogenous contrast agents frequently to enhance the contrast from soft tissues. The agents' cost and potential for side effects and allergic reactions (most of the agents are radioactive, albeit at low levels) limits the utility of the technology. In this context, *quantitative ultrasound* (QUS) methods [6], [9] provide a promising framework that can noninvasively, inexpensively, and quickly be used to assess tumor response to cancer treatment using standard clinical ultrasound equipment [10]. Unlike other aforementioned imaging modalities used in monitoring cancer therapy effects, QUS does not

need injection of any contrast agent. This is mainly because cancer therapies, such as chemotherapy, radiotherapy, or photodynamic therapy, generally aim at inducing programmed cell death, i.e., apoptosis [11], in which many morphological changes occur in tumor cells that affect the bioacoustic properties of tumor. These include nuclear condensation and fragmentation, cell swelling, and chromatin dissolution, which have been demonstrated to directly or indirectly affect the ultrasound backscatter characteristics of tumor [9], [10], [12]. For example, early cell-death (nucleus condensation and fragmentation) increases the backscatter intensity. On the other hand, late cell-death (chromatin dissolution/nucleus degeneration due to apoptosis) has two opposite effects on ultrasound backscatter intensity. This initially increases the randomness of fairly regular backscatter distributions of micro-echoes, which produces a large backscatter signal. However, when a large fraction of the nuclei (scatterers of ultrasound) become disintegrated due to DNA cleavage (advanced necrosis), the amplitude of the backscattered signal has been observed to be reduced [13], [14], [15]. QUS has the main advantage over conventional ultrasound imaging as it uses metrics that are predominantly independent of the instrument settings. Such quantitative measures are derived via analyzing the radio-frequency (RF) raw data before forming B-mode images, i.e., before detecting its envelope and log amplification, and describe bio-acoustic characteristics of the scanned tissue. QUS methods often involve the spectrum analysis (spectroscopy) of the backscatter RF signals over a region of interest, forming *parametric maps*.

The applicability of QUS spectroscopy to detect, noninvasively, cell death has been demonstrated recently *in vitro*, *in situ*, and *in vivo* in both preclinical models and clinical settings using high- and conventional-frequency ultrasound [6], [8], [10], [12], [16]. The initial research to utilize ultrasound to detect apoptosis resulting from cancer treatment *in vitro* and then *in vivo* were reported in [17] and [13], which used high-frequency ultrasound (20–50 MHz). This has recently been extended to conventional clinical US ranges (1–50 MHz) [10], [12], [16], [18], which enables much broader adoption of the technology in the clinic. It builds on the earlier theoretical work of [19], the seminal work behind QUS spectroscopic methods, that utilized spectrum analysis of the conventional low-frequency (5–15 MHz) RF signal to characterize tissues at much coarser scales, e.g., to distinguish between normal and detached retinal tissue.

While previous work in the field of QUS evaluation of cell death response mainly focused on analyses such as statistical tests of significance, regression, and discriminant analysis in order to demonstrate the proof of principle of the proposed QUS systems [9], [18], [16], [10], a next step towards the design of a complete computer-aided-prognosis system is to deploy advanced machine learning algorithms and supervised learning techniques to reliably classify the level of cell death in an automatic manner. Categorizing the extent of the cell death (low, mid, high), noninvasively, is important in the application of monitoring cancer therapy effects. This can be achieved with the settings applied in this study by classifying the extent of cell death to less/more than 20% and 40%.

Relatively few studies have attempted to use supervised learning to detect tumor response to treatment using medical imaging data. Larkin *et al.* [20] worked with $T_2$-weighted MRI to detect cell death, using the support vector machine (SVM) classifier [21] and a "Minkowski functional" feature to classify whether an image represented a treated tumor, or a nontreated control image. Classification accuracy after a 24-h period was 75% using 19 subjects in preclinical data. Histological analysis was used to confirm significant increases in cell death after treatment. SVM was also applied with the goal of differentiating between "pre-" and "post-treatment" images of preclinical data using QUS in [22]; classification accuracy was 87.3% when assessed 24 h after treatment. Two studies from a related field, computer-aided pathology detection, may also suggest appropriate supervised learning methods. Sørensen *et al.* [23] used a joint rotation-invariant local binary pattern (LBP) and intensity histogram in a dissimilarity-based classification approach to predict emphysema using CT imaging. The simple *k*-nearest-neighbor (*k*-NN) classifier and Euclidean distance were used, and the classifier's posterior probabilities were deployed to compute the correlation to the ground truth. Feleppa *et al.* [24] used a multi-layer perceptron classifier to distinguish between cancerous and noncancerous tissues of the prostate at a pixel-by-pixel level using QUS mid-band fit and 0-MHz intercept parametric maps (related to ultrasound backscatter intensity and effective acoustic scatterer concentration, respectively). They found SVMs to give a very similar performance.

In this research, steps were taken towards the development of a computer-aided-prognosis system to assess cancer therapy effects, noninvasively, using conventional-frequency QUS spectral parametric maps. Ultrasound data were acquired from xenograft tumor-bearing mice before and at different times after chemotherapy exposure. After computation of intensity histograms of the parametric maps as a feature descriptor, the distances between "pre-" and "post-treatment" samples were computed using a recently introduced kernel-based metric called maximum mean discrepancy (MMD) [25], [26], [27], [28].

MMD is a distance measure, appropriate when there are multiple data samples available from two populations to be compared. It is a kernel-based measure, implying that its computation is reliant on inner products taken in a reproducing kernel Hilbert space (RKHS). By using a kernel function to nonlinearly transform input vectors into a different, possibly higher-dimensional feature space, and computing the population means in this new space, enhanced group separability (compared to, e.g., Euclidean distance in the original feature space) is ideally obtained.

The main contributions of this study are as follows.

1) Proposing a complete computer-aided-prognosis system using MMD and QUS spectroscopy. The system can provide an early indication of cells undergoing apoptosis (programmed cell death) within hours after treatment, using the QUS spectral parametric maps.
2) Providing one of the first studies to apply machine learning techniques such as kernel-based methods and supervised learning on QUS imaging.

3) Introducing MMD as a kernel-based dissimilarity measure, which is a metric, to represent the levels of cell death as a measure of distance between "pre-" and "post-treatment" images of a subject.

4) While MMD has been also used previously in medical imaging, such as in diffusion-weighted MRI [29], it was used as a biomarker feature selection method, i.e., statistical test, using two empirical probability distribution function (PDF) vectors as features. MMD has been also applied previously, in applications other than medical imaging, as a two-sample statistical test, cost function, or a feature selection method. However, to the best of our knowledge, this is the first time that MMD is proposed as a feature in a dissimilarity-based, supervised pattern recognition framework for classification. The $k$-NN classifier has been used in this dissimilarity framework directly on the MMD as features.

Using the intensity histograms of the parametric maps as a feature descriptor, we compared the proposed MMD-based method with three alternate feature representations that use either the $\ell_2$ norm or the difference of magnitude distance. These feature representations are found in the QUS literature for detecting treatment response, and consist of a *texton* texture representation ("Texton") [22], the same intensity histogram used in the proposed MMD approach ("IntHist"), and a representation of each parametric map with its mean intensity ("MeanInt") [9], [16], [30].

We note that fairly little comparative analysis appears to have been reported on the possible feature representations/transformations, similarity measures, and supervised classifiers that are essential components of computer-aided-prognosis systems using QUS technology, which is on the road to commercialization and clinical use. This work aims to take a step forward in this direction.

## II. METHODS

### A. Data Collection and Preparation

In this study, experiments were carried out using 17 severe combined immunodeficiency disease (SCID) mice. One hind leg of each animal was injected with human breast cancer cells (MDA-MB-231), where they were permitted to grow to 7–9 mm sized xenograft tumors. All animals were anaesthetized before imaging using 100 mg/kg ketamine, 5 mg/kg xylazine, and 1 mg/kg acepromazine (CDMV, St. Hyacinthe, QC, Canada) and grouped into five categories. One set of animals remained untreated (control) while all other animals were given chemotherapy treatment using paclitaxel-doxorubicin (150 $mg/m^2$ and 50 $mg/m^2$, respectively) by way of intravenous tail vein injection. Each group was evaluated at a different time, i.e., 0, 4, 12, 24, and 48 h, after chemotherapy exposure (labeled as CONT, 4H, 12H, 24H, and 48H, respectively).

Ultrasound data were acquired from the whole tumor area in each animal before treatment ("pre-treatment" imaging), and at its group-specified time after exposure ("post-treatment" imaging).

Ultrasound RF data were collected using a Sonix RP ultrasound system (Ultrasonix, Vancouver, BC, Canada), applying an L14–5/38 linear transducer with a center frequency of $\sim$7 MHz, focal depth of 1.5 cm, and sampling rate of 40 MHz. The system was used to collect 3-D data with a scan plane separation of $\sim 0.5$ mm with 10–16 scan frames per tumor depending on tumor size.

### B. Histological Analysis (Ground Truth)

Animals were killed immediately after "post-treatment" imaging in order to obtain a histological ground truth for the extent of cell death, which is required for assessing the effectiveness of the proposed method. Analysis was performed on excised tumor samples fixed for 24–48 h in 5% formalin. Hematoxylin and eosin (H&E) staining was carried out on three representative fixed 5 $\mu m$ thick tumor sections. Moreover, *in situ* end nick labeling (ISEL) immunohistochemistry was performed for cell death. Stained slices were imaged using a Leica DC100 microscope with $20\times$ magnification and a Leica DC100 camera connected to a 2 GHz PC equipped with Leica IM1000 software (Leica GmbH, Wetzlar, Germany). Representative microscopic images of ISEL-stained tumors obtained at different times after exposure are shown in Fig. 1. Cell death areas were quantified macroscopically from immunohistochemistry-stained tumor sections using Image-J (NIH, Bethesda, MD, USA) to detect ISEL positive areas, and subsequently averaged over the three representative sections of tumor. Apoptotic cells were also counted manually at higher magnifications ($40\times$) by identifying typical apoptotic bodies. This ground truth value is referred to as the "histological cell-death fraction" in presented results.

### C. Quantitative Ultrasound Analysis

Standard techniques of spectral estimation [31] were used to derive an estimate of the frequency spectrum. To remove the effects of ultrasound beam diffraction and system transfer function [32], the sample (tissue) power spectrum was normalized with the average power spectrum from an agar-embedded glass bead phantom model [33], [34], scanned with the same settings used for the tissue scans.

*1) Primary Features:* The analysis of the ultrasound data was performed across 10 to 16 scan planes typically with a size of $3.8\times3.0$ cm. Standardized regions of interest (ROIs) were extracted for further analysis from tumor centers, which were consistently positioned at the transducer focal depth and within the transducer's depth of field. Applying a sliding window approach on a pixel by pixel basis, the Fourier transform of the RF data was calculated for each scan line, which was windowed using a Hamming function, with a time-bandwidth product of approximately seven, and subsequently averaged to obtain a power spectrum estimate for the sliding window. Each window's power spectrum was consequently normalized using a power spectrum obtained from the reference phantom following the same scan setting.

Linear regression analysis was performed on the calculated normalized power spectrum [35], [36], within a $-6$ dB bandwidth from the transducer's center frequency (4.5–9.0 MHz), yielding three parameters, which we shall refer to as *primary*
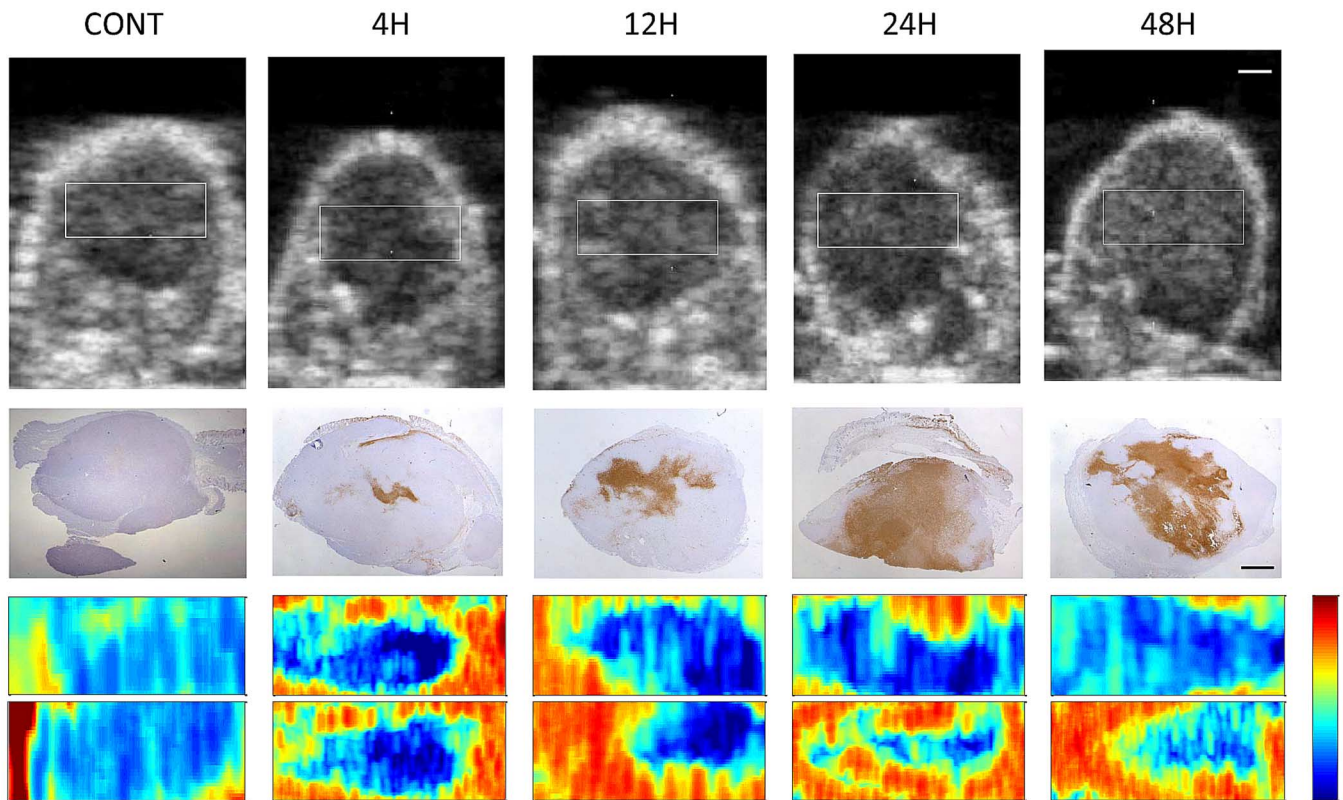
Fig. 1. Representative B-mode Images (first row), the low-magnification microscopic images of ISEL-stained tumors (second row), as well as "pre-treatment" (third row) and "post-treatment" (fourth row) QUS parametric maps of the mid-band fit parameter for each group assessed at a different time after chemotherapy exposure. Rectangles on the B-mode scans indicate where the ROIs are taken from. The color bar represents a scale encompassing approximately 35 dBr and the scale bar represents $\sim$ 1 mm.

*features*: 1) the intercept of the fit line to the calibrated y-axis, termed the *0-MHz intercept* (or intercept in short); 2) the slope of the fit line, termed the *spectral slope* (SS), and 3) the *mid-band fit* (MBF), the normalized power (in dBr) at the center frequency $f_c$. These quantitative parameters can be related to acoustic concentration, effective scatterer size, and ultrasound backscatter power, respectively [35], [36]. The parameters were extracted at each position of the sliding window, thus forming three *parametric maps*. Typical "pre-" and "post-treatment" MBF parametric maps are shown in Fig. 1 over different exposure time intervals.

Previous studies applied different sets of QUS parameters for tissue characterization [24], [37], [38], [39] and therapy response monitoring [9], [18], [22], [30], [40]. In this study, 0-MHz intercept, SS, and MBF were extracted. However, since SS did not show statistically significant changes between pre- and post-treatment ROIs, it was not reported here.

*2) Feature Transformations:* Several feature transformations have been introduced that attempt to reduce the region of interest (ROI) dimensionality or to represent it in a more discriminative fashion. Probability density estimates, used as features, can be effective data descriptors for a large, multidimensional object, such as an image. Due to the very high intrinsic dimensionality and unconstrained input environments of these domains, we believe it is inappropriate to assume a distribution on the data. Alternatively, nonparametric density estimates of the features can be used to represent the data. No assumptions

are made about an underlying distribution of the data (whether it be Gaussian, Poisson, etc.). Nonparametric models support multimodal data, without the complexities of estimating mixture models at the cost of being frequently higher-dimensional than a parametric model.

Traditionally, the mean of the parametric maps are frequently used in the literature to represent them [9], [30]. While this may perform well in some applications, it ignores the rich information contained in parametric maps. Recently, several researchers have proposed treating parametric maps as images, and have applied texture analysis on them. For example, a texton-based approach was applied by Gangeh *et al.* to distinguish between "pre-" and "post-treatment" images [22]. Texton-based approach learns a dictionary of texture atoms (textons) and subsequently represent each ROI by a histogram of these textons [41], [42], [43], [44].

In this study, each ROI was represented with the intensity histograms (as a nonparametric density estimator) of the two corresponding parametric maps. A comparison of each histogram-based parameter with its mean of the parametric map as well as with the texton-based approach is reported in the results section.

### D. Maximum Mean Discrepancy

Maximum mean discrepancy (MMD) is a kernel-based modern approach that addresses the problem of comparing the data samples from two probability distributions. The motivation of using MMD in this study is two fold. 1) It is assumed that the

parametric maps taken from "pre-" and "post-treatment" come from two distributions and the distance between these two distributions can be computed using MMD. If the distance is far, it indicates that the treatment has been effective and the data samples are indeed from two different distributions. However, if they are close, it means that the parametric maps from "pre-" and "post-treatment" are not very much different, thus, the treatment has been ineffective [10]. 2) By using a kernel-based approach such as MMD instead of simple measures like the $\ell_2$ norm, we benefit from mapping data to a high (possibly infinite) dimensional feature space where computing sample means in this space takes into account high order statistics (up to infinity), hence, better discrimination can be achieved.

The concept of MMD [25] is based on Müller's definition of an *integral probability metric* [45]. This metric was designed as a measure to compare the dissimilarity of probability measures[1] $P, Q$, and depends on finding a function $f$ from amongst the space of functions $\mathcal{F}$ that can maximize the distance

$$d_{\mathcal{F}}(P,Q) := \sup_{f \in \mathcal{F}} \left| \int f dP - \int f dQ \right|. \tag{1}$$

If $x \sim P$ and $y \sim Q$, i.e., $x$ and $y$ have distributions $P$ and $Q$, respectively, (1) can be written as

$$\mathbf{MMD}(\mathcal{F},P,Q) := \sup_{f \in \mathcal{F}} (\mathbf{E}_P[f(x)] - \mathbf{E}_Q[f(y)]) \tag{2}$$

where $\mathbf{E}$ is the expectation function. If it is assumed that $\mathcal{F}$ is a unit ball in RKHS $\mathcal{H}$, it can be proved that $\mathbf{MMD}(\mathcal{F},P,Q) = 0$, if and only if $P = Q$ [28], a nice property which is very useful in our application. Hence, from now on, the definition of MMD is restricted on the class of functions $\mathcal{F}$ as the unit ball in RKHS $\mathcal{H}$.

Since $\mathcal{H}$ is an RKHS, according to the Riesz representation theorem [46], there is a feature mapping $\phi(x) : \mathcal{X} \rightarrow \mathbb{R}$ such that $f(x) = \langle f, \phi(x) \rangle_{\mathcal{H}}$ and $k(x,x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}} = \phi(x)^{\top} \phi(x')$, where $\langle .,. \rangle$ is the inner product operator and $k(x,x')$ is a positive definite kernel function. Based on this, (2) can be written as [26], [28]

$$\mathbf{MMD}(\mathcal{F},P,Q) = \| \mathbf{E}_P[\phi(x)] - \mathbf{E}_Q[\phi(y)] \|_{\mathcal{H}}. \tag{3}$$

By working with the squared version of the MMD and knowing that $\|x\|_{\mathcal{H}}^2 = \langle x,x' \rangle_{\mathcal{H}}$, we are able to identify an elegant method of computing it using kernels

$$\begin{aligned}
&\mathbf{MMD}^2(\mathcal{F},P,Q) \\
&= \| \mathbf{E}_P[\phi(x)] - \mathbf{E}_Q[\phi(y)] \|_{\mathcal{H}}^2 \\
&= \langle \mathbf{E}_P[\phi(x)] - \mathbf{E}_Q[\phi(y)], \mathbf{E}_P[\phi(x\prime)] - \mathbf{E}_Q[\phi(y\prime)] \rangle_{\mathcal{H}} \\
&= \mathbf{E}_P \langle \phi(x), \phi(x\prime) \rangle_{\mathcal{H}} - 2\mathbf{E}_{P,Q} \langle \phi(x), \phi(y) \rangle_{\mathcal{H}} \\
&\quad + \mathbf{E}_Q \langle \phi(y), \phi(y\prime) \rangle_{\mathcal{H}} \\
&= \mathbf{E}_{x,x\prime}[k(x,x\prime)] - 2\mathbf{E}_{x,y}[k(x,y)] + \mathbf{E}_{y,y\prime}[k(y,y\prime)] \tag{4}
\end{aligned}$$

where $x,x\prime \sim P$ and $y,y\prime \sim Q$.

In practice, (4) is computed empirically using finite number of data samples taken from the distributions $P$ and $Q$. To this end, if $X = \{x_i\}_{i=1}^n$ and $Y = \{y_i\}_{i=1}^m$ are $n$ and $m$ data samples drawn independently and identically distributed (i.i.d) from $P$

and $Q$, respectively, the empirical $\mathbf{MMD}^2$ can be computed using

$$\mathbf{MMD}^2(\mathcal{F},X,Y) = \frac{1}{n^2} \sum_{i,j=1}^n k(x_i,x_j) - \frac{2}{nm} \sum_{i,j=1}^{n,m} k(x_i,y_j) \\ + \frac{1}{m^2} \sum_{i,j=1}^m k(y_i,y_j). \tag{5}$$

As can be seen from (4) and (5), the data is mapped from the original feature space to the RKHS using a *kernel* $k(x_i,x_j)$, a positive definite function which can perform nonlinear transformations on the data, thereby potentially enhancing the accuracy of linear discriminants in this alternate feature space. Empirical MMD can be computed efficiently for $n + m$ data samples in $\mathcal{O}((n + m)^2)$ time[2], and therefore compared to other dissimilarity measures for distributions such as Parzen estimation or divergences, MMD is far more suited to real-time image analysis.

While MMD was originally proposed in the literature as a technique for statistical hypothesis test [25], [26], [28], it is proposed here as a feature to represent the distance between "pre-" and "post-treatment" data samples of each subject in a dissimilarity space. This enables using computed MMD values in a supervised learning approach, e.g., by submitting them as features to a classifier. Moreover, this provides the possibility for combining them with other features (if needed) to enhance discrimination power of the designed classification system. It is expected that considering MMD as a dissimilarity measure to be useful in exploiting intra-group variance information available from multiple samples/instances taken of each of the "pre-" or "post-treatment" populations. This view on MMD values has been already proved to be useful in scene change point detection in an *unsupervised learning* paradigm [47].

### E. Proposed Computer-Aided-Prognosis System

After the parametric maps have been prepared for each ROI, the 0-MHz intercept and MBF values for each parametric map are separately gathered into normalized (to unit area) histograms of intensity values. These uniformly spaced histograms have $B$ bins, and constitute a rotation- and scale-invariant nonparametric density estimate of the feature.

The MMD distance between "pre-" and "post-treatment" samples of each subject (animal) is then computed using the empirical formulation given in (5). The first term in (5) computes the similarities among all $n$ "pre-treatment" ROIs of each subject whereas the last term computes the same for all $m$ "post-treatment" ROIs of the same subject. The middle term, on the other hand, computes the $n \times m$ pairwise similarities between the $n$ ROIs in the "pre-treatment" and the $m$ ROIs in the "post-treatment" of the same subject. The histogram intersection kernel (HIK) has been selected as the kernel in these computations with the formulation

$$k_{\mathbf{HIK}}(H_1, H_2) = \sum_{b=1}^B \min (h_{1b}, h_{2b}) \tag{6}$$

---

[1] A probability measure has unit area.

[2] An alternative approach has been proposed in [28] with a computational cost of $\mathcal{O}(n+m)$, which is useful in case of existing large data samples but it is not the case in this study.

where $H_1$ and $H_2$ are two intensity histograms with $B$ bins and $h_{1b}$ and $h_{2b}$ values in each bin, respectively. HIK is a parameter-free kernel, which has been used to good effect in many image analysis applications [48].

The last stage is to train a classifier such as a $k$-NN using the MMD values as features, and the ground truth values as labels in a dissimilarity-based classification scheme [49]. This approach, of treating distances or dissimilarities directly as features, requires a symmetric dissimilarity measure, ideally obeying the triangle inequality, such as MMD. The $k$-NN classifier has been used as it is one of the simplest possible classifiers commonly used in dissimilarity space [50]. In this way, attention can be focused on the relative performance differences between MMD and alternatives.

## III. EXPERIMENTAL SETUP AND RESULTS

To set the parameters of the system, a grid search was performed on histograms of sizes 2–200 bins, and the histogram size yielding lowest $k$-NN classification error were selected. This was nine bins for MBF, and 10 bins for the QUS 0-MHz intercept feature. The value of $k$ in the $k$-NN classifier was optimized with respect to the leave-one-out error on the training set.

Two main performance evaluation criteria were selected: the two-sample $t$-test and the ability to predict cell death. The naming scheme used throughout the figures, charts and text is

[QUS Feature]-[Representation]-[Distance Measure]-[Kernel]

e.g., Intercept-IntHist-MMD-HIK.

The slope primary feature performed poorly across all feature representations and evaluation metrics tested, and so its results have been omitted. Development took place on a contemporary Windows Core i5–2520M machine with 4 GB of RAM, using MATLAB (R2011a, MathWorks, Natick, MA, USA).

### A. Alternative Solutions Tested

We compare our proposed approach with three alternative systems published in the recent literature on detecting tumor response changes using QUS.

Gangeh *et al.* [22] proposed treating the parametric images of QUS primary features (e.g., Fig. 1) as textures that can be analyzed using the bag-of-textons approach, which is one of the state-of-the-art texture representation methods. Working with each parametric map in turn, this approach was implemented by extracting 500 randomly chosen patches of size $p \times p$ from each image, and computing the dictionary to form the bag of textons ("Texton") by using $k$-means clustering separately over the set of each subject's "pre-" and "post-treatment" images. The per-subject codebooks[3] were then concatenated together. Thus, if there are $N$ subjects and $k$ atoms per set, the final dictionary is of size $2Nk$. The value of $k = 10$ in $k$-means and patch size of $5 \times 5$ was selected, based on a classification-error-minimizing grid search of $k$ values between 5 and 25 and patch sizes from $3 \times 3$ to $12 \times 12$. The bag-of-textons histogram descriptor was then formed for each ROI using the final codebook.

The second comparison used the same feature representation as our MMD approach, intensity histograms ("IntHist"), but

---

[3]Dictionary, bag of textons, and codebook are used interchangeably in the literature to refer to the same concept.
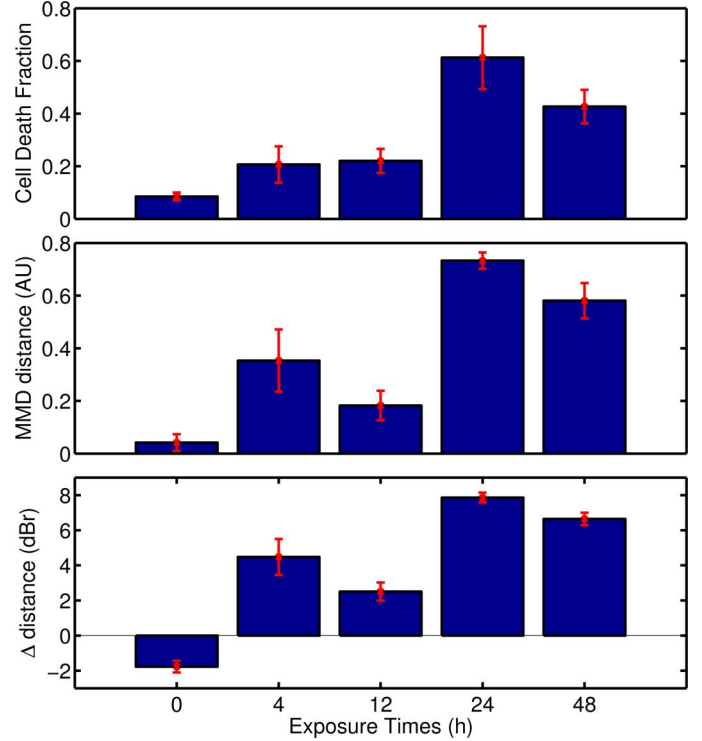


Fig. 2. Cell death, MMD (arbitrary units), and $\Delta$ distance (relative dB, dBr) versus different assessment times after exposure (the MMD and $\Delta$ distances are shown for MBF parametric map). The error bars show the $\pm$ standard error of the mean (SEM).

computed the distance using the Euclidean distance (equivalent to $\ell_2$ norm) instead of MMD.

The third and final comparison was made by representing each ROI using the mean value ("MeanInt"), a scalar value, $\{\mu_{\mathrm{pre}_i}\}_{i=1}^n$ and $\{\mu_{\mathrm{post}_i}\}_{i=1}^m$ of its parametric feature map, and computing the overall means $\mu_{\mathrm{pre}} = 1/n \sum_{i=1}^n \mu_{\mathrm{pre}_i}$ and $\mu_{\mathrm{post}} = 1/m \sum_{i=1}^m \mu_{\mathrm{post}_i}$ separately for each animal. The distance was computed as $\Delta = \mu_{\mathrm{post}} - \mu_{\mathrm{pre}}$ for each subject. This approach was implicitly used in the majority of QUS works that compute 2-D feature plots of primary features, e.g., [30], or report the average "pre-" and "post-treatment" differences of a primary feature, e.g., [9], [16].

Fig. 2 compared the average cell death, Midband-IntHist-MMD-HIK distances, and Midband-MeanInt-$\Delta$ distances by exposure time group. Due to the different vertical scales, it is difficult to compare MMD and $\Delta$ directly. In order to quantitatively compare the histograms in Fig. 2, we again turn to a natural measure, the histogram intersection kernel (6). The MMD and $\Delta$ distances are normalized to have the same value as the cell-death fraction at 24H group. Comparisons against the ground truth histogram yielded a similarity score of 1.433 with Midband-IntHist-MMD-HIK, and 1.303 for Midband-MeanInt-$\Delta$ (higher scores indicate higher similarity), a 10% higher similarity value for MMD.

### B. Statistically Significant Differences From Control Group

In order to compare the statistical significance of the different computed distances, Welch's unpaired two-sample $t$-test [51] was performed using a significance level of $\alpha = 0.05$. The

TABLE I
TEST OF SATISTICAL SGNIFICANCE (UPAIRED, TWO-SAMPLE T-TEST). $p$-VALUES SHOWN FOR EACH EXPOSURE GROUP VERSUS. THE CONTROL GROUP. $^*$ DENOTES $p < 0.05$; $^{**}$ DENOTES $p < 0.01$; $^{***}$ DENOTES $p < 0.001$

| Primary Feature | Group | CONT | 4H | 12H | 24H | 48H |
|---|---|---|---|---|---|---|
| | Number of subjects $N$ | 2 | 3 | 5 | 4 | 3 |
| | Percentage of subjects with $> 20\%$ cell death | 0% | 60% | 66% | 100% | 100% |
| | Avg. cell death (%) | $8.5 \pm 2.1$ | $20.7 \pm 11.9$ | $22.0 \pm 10.2$ | $61.3 \pm 23.8$ | $42.7 \pm 11.0$ |
| | Ground truth $p$-values against control | | 0.268 | 0.140 | 0.042* | 0.026* |
| 0-MHz Intercept | Intercept-MeanInt-$\Delta$ | | 0.037* | 0.123 | 0.002** | 0.005** |
| | Intercept-Texton-MMD-HIK | | 0.249 | 0.183 | 0.004** | 0.038* |
| | Intercept-IntHist-MMD-HIK | | 0.257 | 0.325 | 0.004** | 0.022* |
| | Intercept-IntHist-$\ell_2$ | | 0.310 | 0.390 | 0.035* | 0.060 |
| Mid-band Fit | Midband-MeanInt-$\Delta$ | | 0.019* | 0.005** | 0.000*** | 0.000*** |
| | Midband-Texton-MMD-HIK | | 0.108 | 0.017* | 0.000*** | 0.006** |
| | Midband-IntHist-MMD-HIK | | 0.137 | 0.199 | 0.000*** | 0.009** |
| | Midband-IntHist-$\ell_2$ | | 0.150 | 0.280 | 0.006** | 0.003** |

first group consisted of all of the control subjects' distances, while the second group contained a treatment group's distances. Results are reported in Table I for the different feature-distance combinations, and are roughly arranged in order of ascending $p$ values.

Almost all of the feature-distance combinations has a $p < 0.05$ (the standard threshold of significance) after 24 h of treatment, and $p < 0.01$ when using the MBF. The results are most interesting for the (Intercept/Midband)-MeanInt-$\Delta$ combinations which stand almost alone in reporting significant differences for the 4H and 12H groups (the exception is Midband-Texton-MMD-HIK for the 12H group). As a $t$-test between control group and noncontrol group ground truth cell-death fractions, (Table I) indicates that the 4H and 12H groups do not possess statistically significant differences, this suggests that the mean intensity feature and $\Delta$ distance are overly sensitive and are reporting a false positive, in contrast to MMD.

As a practical matter, it is unlikely that a clinician would change treatment protocols after such a short period, when it is known that the peak response is typically seen 24 h after treatment administration (Fig. 2) [30], [40]. In [9], monitoring photodynamic therapy effects were investigated in preclinical data after intervals as short as 1 h. Although they did not report the $p$ values per exposure group, they showed that the maximum changes in MBF occurred 12 h after treatment while the maximum cell death appeared to happen 24 h after treatment.

Statistical power tests were performed in order to evaluate the reliability of performed statistical tests of significance, considering the use of $t$-test with relatively small sample size in each group [52], [53]. Obtained results indicated that there were no principle objection for using the $t$-test in this study since the obtained statistical power was above 80% for the cases where a statistical significance was reported for the proposed method.

### C. Predicting Cell Death Over a Threshold

Next, a supervised binary classifier was trained, using the distance value between "pre-" and "post-treatment" populations as the sole feature, and cell-death fractions as ground truth. The target is to predict whether or not a subject will have cell death greater than a threshold $\tau$. Ten-fold cross-validation at subject level was performed with the $k$-nearest neighbors ($k$-NN) classifier to successively test the instances. The $k$-NN is the most

common choice in dissimilarity representation [49], which is able to represent complex, multimodal classification surfaces.

Classification was performed using two cell death thresholds, 20% and 40%. Based on the distribution of cell deaths in the population, noticeable gaps around the 20% and 40% levels have been observed, and therefore it was hypothesized that these levels were less susceptible to misclassifications caused by noise in the feature values. The classification accuracy, area under curve (AUC) of the receiver-operator curve (ROC), and Type I (1—Sensitivity) and Type II (1—Specificity) error rates are summarized in Tables II and III for the 20% and 40% thresholds, respectively. Entries are sorted in order of ascending test error. Class priors were set to their observed frequencies, and class-weighted classification error results are reported.

Generally, the MBF feature proved more discriminative in terms of classification error compared to the 0-MHz intercept. MMD with the histogram of MBF values had the lowest error at both threshold levels.

### D. Discussion

The results of this paper demonstrated the viability of the proposed computer-aided-prognosis system using QUS methods at clinically relevant conventional frequencies in conjunction with advanced machine learning techniques in supervised learning paradigm. The proposed system can progressively monitor and categorize the level of cell death with high accuracy and noninvasively on xenograft tumor models *in vivo*.

Overall, the Midband-IntHist-MMD-HIK combination had the strongest performance amongst the different feature-distance combinations on the evaluated metrics and thresholds. Fig. 3 visually compares the MMD versus the Euclidean distances in dissimilarity space. With MMD, we can observe that low cell death subjects are clustered more tightly into a corner compared to $\ell_2$ norm, and the inter-class distance is increased. This reduces the likelihood of $k$-NN errors as well as the errors of classifiers employing linear discriminant functions. Similar scatter plots are obtained if each primary feature, i.e., MBF or intercept, is plotted separately that explains why a $k$-NN classifier performs better using MMD as dissimilarity measure compared with $\ell_2$ norm. Combinations of features in early feature fusion configurations, as well as late classifier-level

TABLE II
CLASSIFICATION ERROR PERCENTAGES AND $\pm 1 \, \sigma$, PREDICTING WHETHER CELL DEATH IS LESS/MORE THAN $\tau = 20\%$. THE $k$ PARAMETER IN $k$-NN CLASSIFIER WAS OPTIMIZED BY LEAVE-ONE-SUBJECT-OUT ON THE TRAINING SET. TEN-FOLD CROSS-VALIDATION AT SUBJECT LEVEL WAS USED; RESULTS SHOWN ARE AVERAGED OVER TEN RUNS. AREA UNDER (THE RECEIVER-OPERATOR) CURVE (AUC) AND TYPE I/II ERROR ARE PROVIDED

| Method | $k$-NN (20%) error | ROC AUC (%) | Type I error (%) | Type II error (%) |
|---|---|---|---|---|
| Midband-IntHist-MMD-HIK | **0.153 ± 0.030** | **0.877** | 0.142 | 0.180 |
| Intercept-IntHist-MMD-HIK | 0.177 ± 0.000 | 0.812 | 0.250 | **0.000** |
| Midband-MeanInt-$\Delta$ | 0.182 ± 0.033 | 0.730 | 0.167 | 0.220 |
| Midband-Texton-MMD-HIK | 0.218 ± 0.028 | 0.831 | 0.158 | 0.360 |
| Intercept-Texton-MMD-HIK | 0.224 ± 0.037 | 0.749 | 0.308 | 0.020 |
| Midband-IntHist-$\ell_2$ | 0.259 ± 0.030 | 0.755 | 0.242 | 0.300 |
| Intercept-MeanInt-$\Delta$ | 0.265 ± 0.042 | 0.825 | 0.225 | 0.360 |
| Intercept-IntHist-$\ell_2$ | 0.294 ± 0.039 | 0.805 | 0.250 | 0.400 |

TABLE III
CLASSIFICATION ERROR PERCENTAGES AND $\pm 1 \, \sigma$, PREDICTING WHETHER CELL DEATH IS LESS/MORE THAN $\tau = 40\%$. THE $k$ PARAMETER IN $k$-NN CLASSIFIER WAS OPTIMIZED BY LEAVE-ONE-SUBJECT-OUT ON THE TRAINING SET. TEN-FOLD CROSS-VALIDATION AT SUBJECT LEVEL WAS USED; RESULTS SHOWN ARE AVERAGED OVER TEN RUNS. AREA UNDER (THE RECEIVER-OPERATOR) CURVE (AUC) AND TYPE I/II ERROR ARE PROVIDED

| Method | $k$-NN (40%) error | ROC AUC (%) | Type I error (%) | Type II error (%) |
|---|---|---|---|---|
| Midband-IntHist-MMD-HIK | **0.118** ± 0.000 | 0.770 | 0.200 | **0.083** |
| Midband-IntHist-$\ell_2$ | 0.229 ± 0.033 | 0.858 | **0.180** | 0.250 |
| Intercept-IntHist-$\ell_2$ | 0.229 ± 0.019 | 0.807 | 0.200 | 0.242 |
| Midband-MeanInt-$\Delta$ | 0.235 ± 0.028 | 0.663 | 0.240 | 0.233 |
| Midband-Texton-MMD-HIK | 0.288 ± 0.019 | **0.882** | 0.400 | 0.242 |
| Intercept-IntHist-MMD-HIK | 0.329 ± 0.050 | 0.757 | 0.520 | 0.250 |
| Intercept-Texton-MMD-HIK | 0.377 ± 0.057 | 0.741 | 0.680 | 0.250 |
| Intercept-MeanInt-$\Delta$ | 0.429 ± 0.056 | 0.684 | 0.660 | 0.333 |

fusion (using Intercept-IntHist-MMD-HIK together with Midband-IntHist-MMD-HIK, for example), slightly improves classification error over using individual features especially at the threshold $\tau = 20\%$, which can be explained using the graphs shown in Fig. 3. As can be seen from the right scatter plots in Fig. 3, especially top right graph, the two classes can be well separated in the space of two primary features using MMD as the dissimilarity measure.

Obtained results implied that changes in the spectral slope parametric maps in response to treatment were less prominent, compared to the MBF and 0-MHz intercept, and were not shown to be statistically significant. In addition to small scattering structures (ensemble of few cells/nuclei undergoing apoptosis), larger scattering structures, such as patches of response and developed gland-like features in MDA human breast tumors, could also affect the spectral slope in an opposite manner. As a working hypothesis, nonsignificant change of slope suggest that both small and large scattering structures potentially play a role at these frequencies [10]. Specifically, such different sized scattering structures can possibly cancel out their effects on slope (one increasing, the other decreasing slope), resulting in nonsignificant overall slope alterations.

Intensity histograms outperformed the texton-based approach in the study here. One explanation is to note that the texton-based approach is a dictionary learning approach, which means that the signals/data are represented using a few textons from a usually overcomplete set of *learned* dictionary elements. The
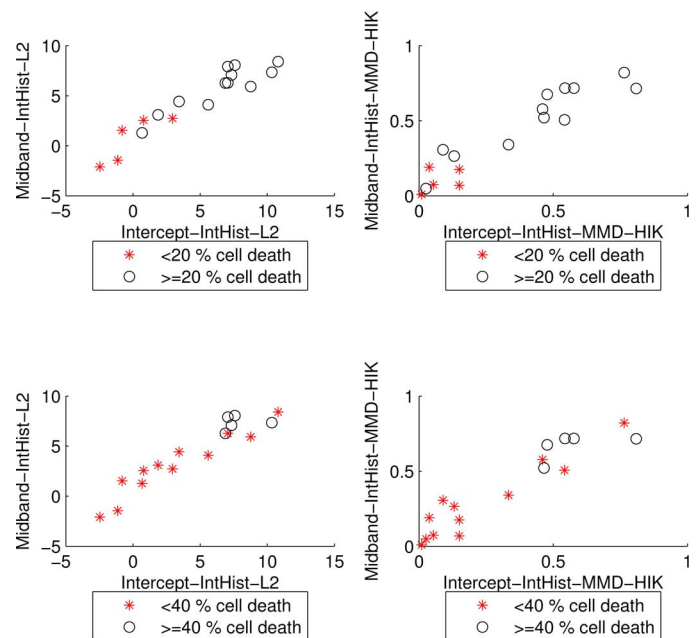


Fig. 3. Scatter plots of the xenograft tumors with various levels of histological cell death presented over the different distance feature planes of MBF and 0-MHZ intercept. $\ell_2$ distance is on left, MMD on right.

learned dictionary can only be representative when there are sufficiently large data samples to train the dictionary. When there

are limited number of training samples, which is the case in this study, learned dictionary may not be sufficiently representative of the data and using some *predefined* basis (dictionary atoms) may lead to a better representation of the underlying distribution and consequently yield better performance in the designed system. This was particularly confirmed in another study on the classification of lung parenchyma in computed tomography (CT) images where, for example, local binary patterns (LBPs) as *predefined* operators outperformed texton-based approach in small sample sizes [54]. Based on this explanation, intensity histogram, which does not need any learning step for a dictionary to represent the underlying data distributions, is expected to outperform a data-driven method like texton-based approach. When more data samples are available for the estimation of the dictionary, the texton-based approach may surpass other approaches in performance.

In this paper, the histogram intersection kernel (HIK) was used for the computation of dissimilarities using MMD. The main advantage of HIK compared with other kernels is that it is parameter free and thus, no tuning has to be performed. However, other kernels such as the radial basis function (RBF) kernel can also be used. The RBF kernel has the main advantage of mapping the data to an infinite dimensional feature space, where the computation of the MMD can take into account the statistics of the data samples up to infinity order [27]. Although the RBF kernel has good analytical properties [55] and is default kernel for generic data [56], it usually performs poorly on histogram descriptors [57] used frequently in image analysis and computer vision. One critical issue in using the RBF kernel is the proper tuning of its parameter that has a great impact on the performance of the systems designed using MMD as feature. This is especially a critical issue when there is a limited number of data samples as the tuning of the RBF kernel may lead to overfitting and hence, degrade the generalization of the designed system. The HIK, can prevent overfitting in this situation. The HIK has also been recommended as the kernel of choice in other works [48], [58] especially on histogram descriptors.

In this study, using the MMD distance as a feature, statistically significant differences were detected using the unpaired *t*-test between treated animals and untreated controls within 24 h after treatment administration. Moreover, it was demonstrated that using the proposed method, cell death can be categorized noninvasively to less/more than 20% and 40% with high accuracy, i.e., 84.7% and 88.2%, respectively. At these two thresholds, a sensitivity of 85.8% and 80% was achieved, respectively while the specificity was 82% and 91.7%. This demonstrates the capability of the proposed approach in categorizing cell death to low, mid, and high levels, which is important in monitoring cancer therapy effects.

The results of this study are in line with recent findings indicating the effectiveness of QUS at conventional frequency in monitoring tumor responses to cancer therapy in preclinical models [10], [16] and clinical settings [12]. Also, the technique proposed in this paper can be considered as a complementary approach to other imaging modalities used in monitoring cancer therapy effects such as positron emission tomography and magnetic resonance imaging [5], [7], [59] with the

main advantages of being cheaper and alleviating the requirement for using an external contrast agent. Furthermore, the proposed computer-aided-prognosis system, which was tested on preclinical data, can be easily adopted to clinical settings [12], [60] in a step forward towards personalized medicine.

## IV. CONCLUSION

A computer-aided-prognosis system was presented for noninvasive tumor response assessment using QUS parametric maps, containing several aspects novel to analysis of QUS data: a dissimilarity-based classification scheme employing the MMD distance measure as features, and the use of intensity histograms of primary features. Three alternative, commonly used feature representation and distance schemes were implemented for comparison purposes. While all showed statistically significant differences between "pre-" and "post-treatment" groups, significant improvements in classification accuracy were observed using MMD and intensity histograms, for both the MBF and 0-MHz intercept features. The system has a classification accuracy of 84.7% and 88.2% when predicting cell death at the thresholds of $\tau = 20\%$ and $\tau = 40\%$, respectively.

The techniques utilized in this work can be applied to other treatments and pathologies, not just for tumor response, but for the broader problem of pathology detection or treatment response monitoring using medical imaging. The work here may contribute to one possible path forward for a fast, noninvasive and inexpensive computed aided diagnosis system, which can fuse together other metadata and predictors about the patient, such as age, gender, and family history to assist clinicians. The MMD can be used in an additional context in such a setting, that is, to identify the additional features and metadata that will be statistically discriminative between populations.

## REFERENCES

[1] J. S. Ross, "Cancer biomarkers, companion diagnostics and personalized oncology," *Biomarkers Med.*, vol. 5, no. 3, pp. 277–279, 2011.

[2] E. C. Hayden, "Personalized cancer therapy gets closer," *Nature*, vol. 458, no. 7235, pp. 131–132, 2009.

[3] K. Sikora, "Personalized cancer therapy," *Personal. Med.*, vol. 2, no. 1, pp. 5–9, 2005.

[4] L. Fass, "Imaging and cancer: A review," *Molecular Oncol.*, vol. 2, no. 2, pp. 115–152, Aug. 2008.

[5] K. Brindle, "New approaches for imaging tumour responses to treatment," *Nat. Rev. Cancer*, vol. 8, no. 2, pp. 94–107, Feb. 2008.

[6] G. J. Czarnota and M. C. Kolios, "Ultrasound detection of cell death," *Imag. Med.*, vol. 2, no. 1, pp. 17–28, 2010.

[7] A. Sadeghi-Naini, O. Falou, J. M. Hudson, C. Bailey, P. N. Burns, M. J. Yaffe, G. J. Stanisz, M. C. Kolios, and G. J. Czarnota, "Imaging innovations for cancer therapy response monitoring," *Imag. Med.*, vol. 4, no. 3, pp. 311–327, Jun. 2012.

[8] M. C. Kolios and G. J. Czarnota, "Potential use of ultrasound for the detection of cell changes in cancer treatment," *Future Oncol.*, vol. 5, no. 10, pp. 1527–1532, 2009.

[9] B. Banihashemi, R. Vlad, B. Debeljevic, A. Giles, M. C. Kolios, and G. J. Czarnota, "Ultrasound imaging of apoptosis in tumor response: Novel preclinical monitoring of photodynamic therapy effects," *Cancer Res.*, vol. 68, no. 20, pp. 8590–8596, Oct. 2008.

[10] A. Sadeghi-Naini, N. Papanicolau, O. Falou, H. Tadayyon, J. Lee, J. Zubovits, A. Sadeghian, R. Karshafian, A. Al-Mahrouki, A. Giles, M. C. Kolios, and G. J. Czarnota, "Low-frequency quantitative ultrasound imaging of cell death in vivo," *Med. Phys.*, vol. 40, no. 8, p. 082901(1-13), 2013.

[11] R. Gerl and D. L. Vaux, "Apoptosis in the development and treatment of cancer," *Carcinogenesis*, vol. 26, no. 2, pp. 263–270, Feb. 2005.

[12] A. Sadeghi-Naini, O. Falou, J. Zubovits, R. Dent, S. Verma, M. E. Trudeau, J. F. Boileau, J. Spayne, S. Iradji, E. Sofroni, J. Lee, S. Lemon-Wong, M. J. Yaffe, M. C. Kolios, and G. J. Czarnota, "Quantitative ultrasound evaluation of cell death response in locally advanced breast cancer patients receiving chemotherapy," *Clin. Cancer Res.*, vol. 19, no. 8, pp. 2163–2174, 2013.

[13] G. J. Czarnota, M. C. Kolios, J. Abraham, M. Portnoy, F. P. Ottensmeyer, J. W. Hunt, and M. D. Sherar, "Ultrasound imaging of apoptosis: high-resolution non-invasive monitoring of programmed cell death in vitro, in situ and in vivo," *Br. J. Cancer*, vol. 81, no. 3, pp. 520–527, Oct. 1999.

[14] J. W. Hunt, A. E. Worthington, A. Xuan, M. C. Kolios, G. J. Czarnota, and M. D. Sherar, "A model based upon pseudo regular spacing of cells combined with the randomisation of the nuclei can explain the significant changes in high-frequency ultrasound signals during apoptosis," *Ultrasound Med. Biol.*, vol. 28, no. 2, pp. 217–226, 2002.

[15] V. Rouffiac, C. Bouquet, N. Lassau, P. Opolon, S. Koscielny, P. Peronneau, M. Perricaudet, and A. Roche, "Validation of a new method for quantifying in vivo murine tumor necrosis by sonography," *Invest. Radiol.*, vol. 39, no. 6, pp. 350–356, 2004.

[16] A. Sadeghi-Naini, O. Falou, H. Tadayyon, A. Al-Mahrouki, W. Tran, N. Papanicolau, M. C. Kolios, and G. J. Czarnota, "Conventional frequency ultrasonic biomarkers of cancer treatment response in vivo," *Translat. Oncol.*, vol. 6, no. 3, pp. 234–243, Jun. 2013.

[17] G. J. Czarnota, M. C. Kolios, H. Vaziri, S. Benchimol, F. P. Ottensmeyer, M. D. Sherar, and J. W. Hunt, "Ultrasonic biomicroscopy of viable, dead and apoptotic cells," *Ultrasound Med. Biol.*, vol. 23, no. 6, pp. 961–965, 1997.

[18] A. Sadeghi-Naini, O. Falou, and G. J. Czarnota, "Quantitative ultrasound spectral parametric maps: Early surrogates of cancer treatment response," in *Proc. 2012 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 2672–2675.

[19] F. L. Lizzi, M. Greenebaum, E. J. Feleppa, M. Elbaum, and D. J. Coleman, "Theoretical framework for spectrum analysis in ultrasonic tissue characterization," *J. Acoust. Soc. Am.*, vol. 73, no. 4, pp. 1366–1373, Apr. 1983.

[20] T. J. Larkin, H. C. Canuto, M. I. Kettunen, T. C. Booth, D.-E. Hu, A. S. Krishnan, S. E. Bohndiek, A. A. Neves, C. McLachlan, M. P. Hobson, and K. M. Brindle, Magnetic Resonance in Medicine "Analysis of image heterogeneity using 2-D Minkowski functionals detects tumor responses to treatment," Feb. 2013 [Online]. Available: http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1522–2594

[21] T. Hastie, R. Tibshirani, and J. J. H. Friedman, *The Elements of Statistical Learning*, 1st ed ed.   New York: Springer, 2001.

[22] M. J. Gangeh, A. Sadeghi-Naini, M. S. Kamel, and G. Czarnota, "Assessment of cancer therapy effects using texton-based characterization of quantitative ultrasound parametric images," in *Proc. Int. Symp. Biomed. Imag.: From Nano to Macro*, 2013, pp. 1372–1375.

[23] L. Sørensen, S. B. Shaker, and M. d. Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 559–569, Feb. 2010.

[24] E. J. Feleppa, J. Mamou, C. R. Porter, and J. Machi, "Quantitative ultrasound in cancer imaging," *Seminars Oncol.*, vol. 38, no. 1, pp. 136–150, Feb. 2011.

[25] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola, "A kernel method for the two-sample-problem," in *Adv. in Neural Inform. Processing Syst.*   Cambridge, MA: MIT Press, 2006, vol. 19, pp. 513–520.

[26] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, A kernel method for the two-sample problem Max Planck Inst. Biol. Cybern., Tech. Rep. 157, 2008.

[27] S. Jegelka, A. Gretton, B. Schölkopf, B. K. Sriperumbudur, and U. V. Luxburg, "Generalized clustering via kernel embeddings," in *Proc. 32nd Annu. German Conf. Adv. Artif. Intell.*, 2009, pp. 144–152.

[28] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, pp. 723–773, 2012.

[29] Y. Rathi, J. Malcolm, O. Michailovich, J. Goldstein, L. Seidman, R. W. McCarley, C.-F. Westin, and M. E. Shenton, "Biomarkers for identifying first-episode schizophrenia patients using diffusion weighted imaging," in *Proc. MICCAI*, Jan. 2010, vol. 13, pp. 657–665.

[30] R. M. Vlad, S. Brand, A. Giles, M. C. Kolios, and G. J. Czarnota, "Quantitative ultrasound characterization of responses to radiotherapy in cancer mouse models," *Clin. Cancer Res.*, vol. 15, no. 6, pp. 2067–2075, Mar. 2009.

[31] B. Porat, *Digital processing of random signals: Theory and methods*. Mineola, NY: Dover, 1994.

[32] R. Romijn, J. Thijssen, and G. v. Beuningen, "Estimation of scatterer size from backscattered ultrasound: a simulation study," *IEEE Trans. Ultrason., Ferroelectr. Freq. Control*, vol. 36, no. 6, pp. 593–606, Nov., 1989.

[33] F. Dong, E. L. Madsen, M. C. MacDonald, and J. A. Zagzebski, "Nonlinearity parameter for tissue-mimicking materials," *Ultrasound Med. Biol.*, vol. 25, no. 5, pp. 831–838, 1999.

[34] L. X. Yao, J. A. Zagzebski, and E. L. Madsen, "Backscatter coefficient-measurements using a reference phantom to extract depth-dependent instrumentation factors ," *Ultrason. Imag.*, vol. 12, no. 1, pp. 58–70, 1990.

[35] F. Lizzi, M. Ostromogilsky, E. Feleppa, M. Rorke, and M. Yaremko, "Relationship of ultrasonic spectral parameters to features of tissue microstructure," *IEEE Trans. Ultrason., Ferroelectr. Freq. Control*, vol. 34, no. 3, pp. 319–329, May 1987.

[36] F. L. Lizzi, M. Astor, T. Liu, C. Deng, D. J. Coleman, and R. H. Silverman, "Ultrasonic spectrum analysis for tissue assays and therapy evaluation," *Int. J. Imag. Syst. Technol.*, vol. 8, no. 1, pp. 3–10, 1997.

[37] M. Oelze, W. O'Brien, J. Blue, and J. Zachary, "Differentiation and characterization of rat mammary fibroadenomas and 4T1 mouse carcinomas using quantitative ultrasound imaging," *IEEE Trans. Med. Imag.*, vol. 23, no. 6, pp. 764–771, Jun. 2004.

[38] E. Feleppa, A. Kalisz, J. Sokil-Melgar, F. Lizzi, T. Liu, A. Rosado, M. Shao, W. Fair, Y. Wang, M. Cookson, V. Reuter, and W. Heston, "Typing of prostate tissue by ultrasonic spectrum analysis," *IEEE Trans. Ultrason., Ferroelectr. Freq. Control*, vol. 43, no. 4, pp. 609–619, Jul. 1996.

[39] M. Yang, T. M. Krueger, J. G. Miller, and M. R. Holland, "Characterization of anisotropic myocardial backscatter using spectral slope, intercept and midband fit parameters," *Ultrason. Imag.*, vol. 29, no. 2, pp. 122–134, 2007.

[40] J. Lee, R. Karshafian, N. Papanicolau, A. Giles, M. C. Kolios, and G. J. Czarnota, "Quantitative ultrasound for the monitoring of novel microbubble and ultrasound radiosensitization," *Ultrasound Med. Biol.*, vol. 38, no. 7, pp. 1212–1221, Jul. 2012.

[41] B. Julesz, "Textons, the elements of texture perception, and their interactions," *Nature*, vol. 290, no. 5802, pp. 91–97, 1981.

[42] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *Int. J. Comput. Vis.*, vol. 62, no. 1-2, pp. 61–81, Apr. 2005.

[43] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vis.*, vol. 43, pp. 29–44, Jun. 2001.

[44] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2032–2047, Nov. 2009.

[45] A. Müller, "Integral probability metrics and their generating classes of functions," *Adv. Appl. Probabil.*, pp. 429–443, 1997.

[46] M. Reid and B. Simon, *Methods of Modern Mathematical Physics*. San Diego, CA: Academic, 1980, vol. 1, Funct. Anal..

[47] M. Diu, M. J. Gangeh, and M. S. Kamel, "Unsupervised visual change-point detection using maximum mean discrepancy," in *Proc. 10th Int. Conf. Image Anal. Recognit.*, 2013, pp. 336–345.

[48] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 2169–2178.

[49] E. Pekalska and R. P. W. Duin, "Dissimilarity representations allow for building good classifiers," *Pattern Recognit. Lett.*, vol. 23, no. 8, pp. 943–956, Jun. 2002.

[50] E. Pekalska, P. Paclik, and R. P. W. Duin, "A generalized kernel approach to Dissimilarity-based classification," *J. Mach. Learn. Res.*, vol. 2, no. 2, pp. 175–211, 2002.

[51] H. Motulsky, *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking*, 2nd ed. New York: Oxford Univ. Press, 2010.

[52] S. Siegel and N. J. Castellan, *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill, 1988.

[53] J. C. F. d. Winter, "Using the student's t-test with extremely small sample sizes," *Practical Assessment, Res. Evaluat.*, vol. 18, no. 10, 2013.

[54] L. Sørensen, M. J. Gangeh, S. B. Shaker, and M. d. Bruijne, "Texture classification in pulmonary CT," in *Lung Imaging and Computer Aided Diagnosis*, A. El-Baz and J. S. Sure, Eds. Boca Raton, CA: CRC Press, 2007, pp. 343–367.

[55] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[56] R. Fan, P. Chen, and C. Lin, "Working set selection using the second order information for training SVM," *J. Mach. Learn. Res.*, vol. 6, pp. 1889–1918, 2005.

[57] J. Wu, "A fast dual method for HIK SVM learning," in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 552–565.

[58] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 213–238, 2007.

[59] T. H. Witney and K. M. Brindle, "Imaging tumour cell metabolism using hyperpolarized 13C magnetic resonance spectroscopy," *Biochem. Soc. Translat.*, vol. 38, no. 5, pp. 1220–1224, 2010.

[60] N. R. Jagannathan, M. Kumar, V. Seenu, O. Coshic, S. N. Dwivedi, P. K. Julka, A. Srivastava, and G. K. Rath, "Evaluation of total choline from in-vivo volume localized proton MR spectroscopy and its response to neoadjuvant chemotherapy in locally advanced breast cancer," *Br. J. Cancer*, vol. 84, no. 8, pp. 1016–1022, Apr. 2001.