Bio-inspired Computation Approach for Tumor Growth with Spatial Randomness Analysis of Kidney Cancer Xenograft Pathology Slides *

Aydin Saribudak
Electrical Engineering
The City College of New York
New York, NY 10031
asaribudak@ccny.cuny.edu

Yiyu Dong
Memorial Sloan Kettering
Cancer Center
New York, NY 10065
dongy@mskcc.org

James Hsieh
Memorial Sloan Kettering
Cancer Center
New York, NY 10065
hsiehj@mskcc.org

M. Ümit Uyar
Electrical Engineering
The City College of New York
New York, NY 10031
uyar@ccny.cuny.edu

ABSTRACT

In this paper, we analyze digitized images of Hematoxylin-Eosin (H&E) slides equipped with tumorous tissues from patient derived xenograft models to build our bio-inspired computation method, namely Personalized Relevance Parameterization of Spatial Randomness (PReP-SR). Applying spatial pattern analysis techniques of quadrat counts, kernel estimation and nearest neighbor functions to the images of the H&E samples, slide-specific features are extracted to examine the hypothesis that existence of dependency of nuclei positions possesses information of individual tumor characteristics. These features are then used as inputs to PReP-SR to compute tumor growth parameters for exponential-linear model. Differential evolution algorithms are developed for tumor growth parameter computations, where a candidate vector in a population consists of size selection indices for spatial evaluation and weight coefficients for spatial features and their correlations. Using leave-one-out-cross-validation method, we showed that, for a set of H&E slides from kidney cancer patient derived xenograft models, PReP-SR generates personalized model parameters with an average error rate of 13.58%. The promising results indicate that bio-inspired computation techniques may be useful to construct mathematical models with patient specific growth parameters in clinical systems.

*The initial research used in this work was supported by U.S. Army Communications-Electronics RD&E Center contracts W15P7T-09-CS021 and W15P7T-06-C-P217, and by the National Science Foundation grants ECCS-0421159, CNS-0619577 and IIP-1265265. The contents of this document represent the views of the authors and are not necessarily the official views of, or are endorsed by, the U.S. Government, Department of Defense, Department of the Army or the U.S. Army Communications-Electronics RD&E Center.

Keywords

artificial intelligence, bio-inspired computation, differential evolution, kidney cancer, H&E slide, pathology, exponential linear tumor growth model, spatial pattern analysis

1. INTRODUCTION

The growth of tumor mass in living bodies is a result of the tumorous cell proliferation characterized by the biological processes in cancerous tissues. These processes can be analyzed with functional models based on physiological hypotheses and biochemical reactions [1, 2]. Apart from functional models, tumor growth can also be described with empirical models using mathematical statements of growth behavior without taking into account the basal processes [3]. Exponential-linear model [1] is an empirical method of tumor growth, where the amount of tumor mass increases exponentially up to a threshold weight or volume at which the nature of the growth becomes linear as a result of limitations in space and supplies such as nutrition and oxygen [4].

In this study, we utilize Hematoxylin-Eosin (H&E) stained slides to compute exponential-linear model growth parameters using bio-inspired computation techniques. Equipping the glass slides with diseased tissues dyed with a number of colors is a methodology called H&E staining [5, 6]. For determination of cancer type and its level in clinical settings, pathologists are provided with H&E slides where the nuclei of the cells are stained in blue color and the connective tissues in pink by hematoxylin and eosin, respectively [5]. The slides are visualized under a light microscope to determine the amount of nuclei and their positions.

The spatial pattern of the position of points in an Euclidian plane can be quantified with spatial randomness methods of quadrat counts, kernel estimation and nearest neighbor functions [7, 8, 9]. Applying these spatial pattern analysis techniques to the images of the H&E samples, slide-specific features can be extracted to examine the hypothesis that existence of dependency of nuclei positions possesses information of individual tumor characteristics.

In this paper, using the features extracted from spatial randomness analysis of nuclei positioning, we compute tumor growth parameters specific to each H&E slide generated from kidney cancer patient derived xenograft models. Our parameter computation method is called Personalized Relevance Parameterization of

BICT 2015, December 03-05, New York City, United States

DOI 10.4108/eai.3-12-2015.2262350

Spatial Randomness (PReP-SR), which implements a differential evolution (DE) [10] based artificial intelligence algorithm. In PReP-SR a candidate vector in a population consists of size selection indices for spatial evaluation and weight coefficients for spatial features and their correlations. We show that PReP-SR computes personalized tumor growth parameters for 14 H&E slides with an average error rate of 13.58% using leave one out cross validation (LOOCV) method [11].

Sect. 2 of this paper outlines related work in tumor growth models and spatial pattern analysis techniques. In Sect. 3 image descriptors of H&E slides and construction and evaluation algorithms for our PReP-SR model are introduced. Analytical results are discussed in Sect. 4. Concluding remarks are in Sect. 5.

2. RELATED WORK

Empirical models are defined to examine the proliferation of tumor cells mathematically [12]. Exponential linear tumor growth model [1, 3, 13] is an empirical model characterizing tumor progression with exponential and linear stages, where the growth behavior switches from exponential to linear at a certain weight. The stage of a tumorous tissue can be graded experimentally by H&E staining techniques, where the tissue sections are analyzed under a light microscope [5]. Feature extraction methods are used in analysis of pathology slides [5, 14, 15]. In [14], Orlov et al. studied different feature extraction methods for biological and medical images. The graphical and textural features of the digitized H&E stained slides are examined by Doyle et al. [15] to capture the characteristics of nuclei patterns.

Characteristics of object patterns in a spatial region can be evaluated in terms of randomness of point locations [9]. A number of methods of spatial pattern analysis are studied in literature to identify data features [16]. The methods to extract information from randomness of spatial point patterns are classified as the first and second order properties to evaluate the intensity of the observations and the relationship between point pairs [17]. In [18], Perry et al. presented an analysis for different methods of spatial point patterns to provide suggestions for their interpretations. Baddeley et al. [19] introduced applications of spatial point pattern analysis.

In previous phases of our research [20, 21, 6], we studied construction of personalized methods to compute growth characteristics of tumor mass based on biomarkers such as gene expression values and pathology slides using the artificial intelligence based techniques we develop in our Bio-Inspired Computation Laboratory at Electrical Engineering Department of the City College of New York. In [20], we analyze the patient-specific tumor mass proliferation for breast cancer using the genetic information of individuals whose gene expressions and tumor volume measurements are provided in I-SPY 1 TRIAL database [22] and NBIA database [23], respectively. We computed the transit compartmental PK-PD model parameters in [21] using data provided in literature [24, 25], where chemotherapeutic agents are administered to xenograft mice models under different schedules, following the implantation of the cells from breast cancer cell lines. Clustering is a method to group data sets, where each cluster consists of similar data points [26]. Structure of clusters is quantified by a variety of methods reported in literature [27, 28]. In our earlier study [6], we showed that the data retrieved from the clusters of nuclei positioned on H&E stained slides can be used to compute tumor growth parameters.

In this paper, our goal is to compute the patient-specific tumor growth parameters with the help of H&E slide features generated by spatial randomness methods applied to the pathology samples from mice models injected with kidney cancer tumor cells.

3. METHODOLOGY

3.1 Preclinical Study and H&E Staining

To build a systematic approach for the computation of personalized parameters of exponential-linear model, we study spatial randomness of nuclei positions on H&E stained slides from kidney cancer xenograft models. In this research, patient-derived tumor xenograft models were developed in Memorial Sloan Kettering Cancer Center (MSKCC) using a protocol approved by the Institutional Animal Care and Use Committee (IACUC).

4-6 week old NOD scid gamma (NSG) male mice obtained from Jackson Laboratory were anesthetized by inhalation with an isoflurane vaporizer before the tumor implantation process. Primary tumor tissues obtained from kidney cancer patients were cut into $3x3x3\ mm^3$ fragments and implanted into mice flanks subcutaneously. Carprofen was administered to the mice by intraperitoneal injection within 24 hours after surgery when they were still anesthetized.

To derive the growth characteristics of tumor mass in mice flanks, we measured the tumor volume twice per week for approximately a period of one month. The length (L) and width (W) of tumor mass were measured by caliper and its volume was calculated using the formula $0.5 \cdot L \cdot W^2$. Although the experiments were completed at the end of the one month period, any mouse with more than 2 *cm* tumor mass in length in its flanks was euthanized without waiting for the end of the experiment.

With the completion of clinical measurements, tumor growth curves were plotted with GraphPad Prism software and tumor tissue samples were collected and fixed in 10% formalin for histopathological analysis. Tumor tissues were cut into 5 μm sections with microtome, and baked to use for standard H&E staining.

3.2 Spatial Pattern Analysis of H&E Slides

3.2.1 Spatial Point Process

Spatial point processes are statistical tools for analyzing the observed pattern of objects to assess their spatial characteristics, such as clustered, regularly distributed or located randomly in a study area [17]. The randomness in the distribution of objects is determined in terms of spatial stationarity, which can be expressed as the similarity in the patterns of randomly selected subareas of a region [29]. Spatial randomness of objects is measured as the likelihood of their occupance of any location in a space, and the independence of these positions from each other. The randomness of objects' positions in space is based on the principle that regions with the same size should have the same chance of containing any randomly located object [29].

To extract information from a point pattern in a given space, the identification of relative locations of the points is essential [16]. Randomness of point patterns is assessed with the first and second order properties of a spatial point process, where the former properties are used in terms of intensity of the observations and the latter ones are analyzed on the basis of the relationship between point pairs [17].

3.2.2 First and Second Order Properties

Quadrat count and *kernel estimation* are two typical techniques in examining the first order properties based on intensity of observation. In quadrat count, variation in the intensity of observations of equal-sized rectangles provides an indication for the randomness in the distribution of the objects [16]. Like many area-based metrics, quadrat counts may not be precise enough in evaluating the first order properties of randomness due to their dependency on

quadrat dimensions. For this reason, kernel estimation technique is used together with quadrat count. Kernel estimation is determined based on the locations of objects to examine their density with a kernel function of distance measurement.

In this study, we apply quartic kernel estimation method to the nuclei locations of H&E slides in the assessment of the randomness of their locations. Kernel estimation value [8] of $\lambda_r(p_i)$ for a given object p_i can be calculated as:

$$\lambda_r(p_i) = \frac{1}{r^2} \cdot \sum_{p_j \in P \setminus p_i} k\left(p_i, p_j\right) \tag{1}$$

where r is the threshold radius to specify the closest region for object p_i, P is the set of all objects, and $k\left(p_i, p_j\right)$ is the kernel estimation function for two distinct objects of p_i and p_j. Below, k_Q is formulated for the quartic based version of kernel function [8] for objects p_i and p_j, where $\|p_i - p_j\|$ is the euclidian distance between these objects:

$$k_Q\left(p_i, p_j\right) = \frac{3}{\pi} \cdot \left(1 - \frac{\|p_i - p_j\|^2}{r^2}\right) \tag{2}$$

Second order properties such as the nearest neighbor distances are calculated to explore the spatial dependency among objects [17]. For example, according to second order properties, a regular or a clustered pattern of objects violates the spatial randomness among them, where the patterns are caused by an inhibitive or an attractive interaction of object sets, respectively. The function $K(\upsilon)$ [30, 8] defined as a function of threshold kernel radius υ to analyze second order property of pattern randomness is formulated as follows:

$$K(\upsilon) = \frac{R_P}{n^2} \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{I_\upsilon\left(d_{ij}\right)}{w_{ij}} \tag{3a}$$

$$I_\upsilon\left(d_{ij}\right) = \begin{cases} 1, & if \;\; d_{ij} \leq \upsilon \\ 0, & otherwise \end{cases} \tag{3b}$$

where n is the number of points on the plane, R_P is the area on which the points of set P are located and w_{ij} is the weight value to correct the so-called edge effect [30] between objects i and j for the set of P in an arbitrarily bounded region. Note that, due to the fact that pathology images are bounded with straight edges, we can disregard the edge effects in Eq. (3a) (i.e., $w_{ij} = 1 \;\; \forall\, i, j \in P$). To count the total number of points inside a near neighbor region for each object, an indicator function I_υ is defined, which has a value of 1 for the points inside the region determined with the threshold υ and 0 otherwise.

If a complete spatial randomness exists among objects of a given region, *K function* calculates the expected number of observations as $\pi \cdot \upsilon^2$ [8]. For this case, a special function, called *L function* [9], is defined as a standardized version of *K function* to identify the characteristics of second order properties. *L function*, which can grade the dependence of locations quantitatively, is formulated as:

$$L(\upsilon) = \sqrt{\frac{K(\upsilon)}{\pi}} - \upsilon \tag{4}$$

where in the case of complete spatial randomness, $L(\upsilon)$ becomes 0. If the pattern of the objects is a regular or clustered shape, $L(\upsilon)$ evaluates to a negative or a positive value, respectively.

Once the dimensions of quadrats in quadrat count method and the threshold radius of the region of observation in kernel estimation method are selected, first order spatial pattern analysis techniques can be applied to evaluate the randomness of object distribution. The spatial dependency second order property indicators of *K*- and *L-functions* can also be calculated based on threshold radius. Therefore, to extract information from object distributions, a computationally efficient approach to select the dimension and radius values for a given region is required. In our study, we apply bio-inspired computation methods to select the proper dimension and radius values as explained in Sec. 3.3.

3.2.3 *Spatial Randomness Features of* H&E *Slides*

In our study, we analyze digitized images of H&E slides equipped with tumor tissues from patient derived xenograft models to build our personalized parameterization approach. We apply spatial pattern analysis techniques to the images of the H&E slides to examine and grade the randomness of nuclei locations. Slide-specific features are extracted to evaluate the nuclei positions on glass slides for the amount of information they may contain related to individual tumor characteristics. These features are then used as inputs to our personalized parameterization method of spatial randomness (PREP-SR) to compute the parameters of tumor growth models.

In our PREP-SR method, using bio-inspired computation techniques, we compute the first four moments of the first order properties and the *L-function* of the second order properties. Our algorithm converges toward fitter dimension and radius values of first order properties with the computation of moments and *L-function*. The fitness of the algorithm is defined such that, while DE computes the spatial features, it also identifies the most informative quadrat dimensions and radius sizes.

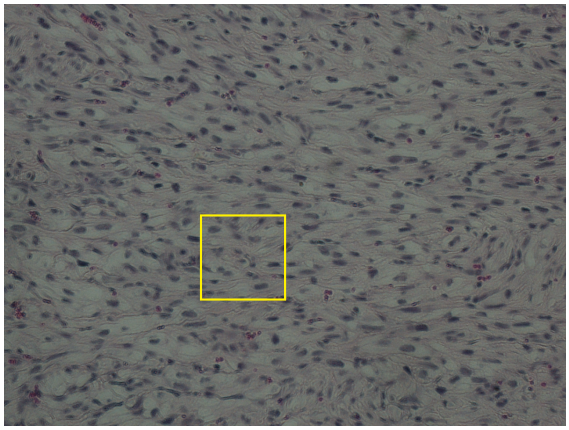
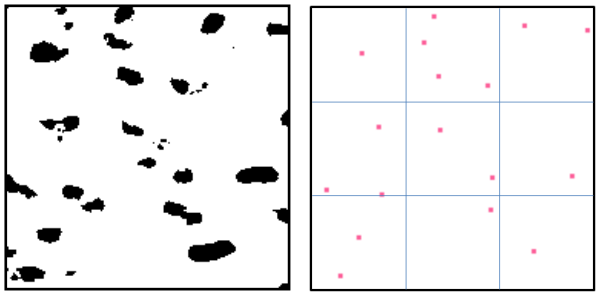


Figure 1: Image of H&E slide of Sample 51013-L digitized under a light microscope with 20x magnification



(a) Binary image region (b) Nuclei centers in quadrats

Figure 2: The binary image of Sample 51013-L for the highlighted part in Fig. 1 and nuclei positions in corresponding quadrats

In H&E staining technique, diseased tissue samples are equipped on glass slides to make the nuclei distribution visible to human eye. As an example of H&E staining, an image of digitized slide is given for Sample 51013-L in Fig. 1, which was taken under a light microscope with 20x magnification. To distinguish the nuclei positions from other particles, we apply image processing techniques such as binary imaging and basic morphology operations.

The binary images generated from the digitized images of pathology slides are processed to extract features using first and second order properties of spatial randomness. Fig. 1a shows the binary image of Sample 51013-L for the yellow box region in Fig. 1, which corresponds to nine quadrats in our calculations. The nuclei locations identified in these quadrats are shown in Fig. 1b. After defining the quadrats on these binary images, the number of nuclei is counted for each quadrat to determine the area-based randomness moments using intensity variation.

In addition to area-based spatial randomness moments, quartic kernel estimation method, which is the location-based method of first order properties, is applied to calculate first four moments of the kernel estimation value (i.e., λ_r) using Eqs. (1) and (2). To generate the set of λ_r values for each nuclei in a slide, their locations are defined as centers of kernel circles. For the second order nuclei distribution characteristics, the functions of the second order properties are computed with Eqs. (3) and (4), where the nuclei locations are defined as nuclei centers similar to the kernel estimation case. The first four moments of the area-based and location-based spatial randomness methods and the resultant *L-function* for the second order property are used as the slide specific features.

In PREP-SR, the dimension and radius values are computed together with spatial pattern features. For the first and second order spatial randomness measures, we compute the fittest values for the solution space determined by the size of the pathology slides. In Table 1, the size and the feature results computed by PREP-SR model for samples 51010-L, 50054, 51013-L and 50052 are listed, where the dimension and radius values are given in the unit of pixel (i.e., *px*), quadrat and kernel feature values in $1/px^2$, and *L-function* in *px*. The slide-specificity of feature values can be seen from the results given in Table 1. For the samples 51010-L and 50054, the values computed by our algorithm for the same features do not show similarity even though the computed size values are close to each other.

3.3 Modeling Tumor Growth

3.3.1 Exponential Linear Tumor Growth Model

Exponential-linear tumor growth model [3, 31] defines parameters specific to the phase of the tumor growth characterized as cell proliferation. In this empirical model, the growth of tumor mass in untreated animal xenograft models has an initial exponential and a followed by a linear behavior [1]. Consistent with clinical observations, tumorous tissue initially starts growing exponentially until the nutrition and space limitations force the growth to switch to a linear behavior [4]. In this model, the tumor weight $w(t)$ at a given time t is formulated as:

$$\frac{dw(t)}{dt} = \begin{cases} \lambda_0 \cdot w(t), & w(t) \leq w_{th} \\ \lambda_1, & w(t) > w_{th} \end{cases} \quad (5)$$

where w_{th} is the threshold weight at which tumor growth characteristics switches from exponential to linear. λ_0 and λ_1 are exponential and linear growth parameters, respectively. The values of these two parameters and w_{th} are determined such that the derivatives in Eq. 5 are kept continuous, which results in $\lambda_1 = \lambda_0 \cdot w_{th}$.

Table 1: Size and Feature values of spatial randomness methods computed for selected H&E slides samples 51010-L, 50054, 51013-L and 50052

Size & Feature	Sample Slide ID			
	51010-L	50054	51013-L	50052
Size Type	Size Values from PREP-SR (*px*)			
Quadrat Width	80	80	64	64
Quadrat Height	60	60	64	80
Kernel Radius	44	43	58	57
NN-circle Radius	54	77	53	41
Feature Name	Feature Values from PREP-SR ($1/px^2$)			
Quadrat Mean	2.20	2.38	2.17	2.77
Quadrat Variance	1.82	1.57	1.32	1.79
Quadrat Skewness	0.33	0.38	0.44	-0.05
Quadrat Kurtosis	2.70	3.02	3.61	2.68
Kernel Mean	2.02	2.09	3.88	3.90
Kernel Variance	1.37	1.38	2.03	2.32
Kernel Skewness	0.42	0.31	0.03	0.34
Kernel Kurtosis	2.95	2.71	2.78	2.97
L-*function* (in *px*)	-3.01	-4.20	-4.43	-4.11

3.3.2 Personalized Tumor Growth Parameters

In our research, we study methodologies to compute personalized parameters for tumor growth models using biomarkers obtained from individuals. We evaluate the clinical relevance of these parameters by using tumor cells from patients and cancer cell lines implanted into xenograft models. For example, in earlier phases of our research [20, 21], we used gene expression values to compute tumor growth parameters of exponential-linear model and the shrinkage parameters of PK-PD compartmental model for breast cancer cell lines. In [6], we utilized H&E slides from patient derived xenograft models to compute the parameters for tumor samples collected from kidney cancer patients. In this paper, we extent our work on H&E slides to identify further image descriptors, which are defined based on nuclei patterns on pathology slides. The level of randomness in nuclei locations is used as a metric to extract quantitative information from slide samples. Using the slide-specific features retrieved with spatial randomness methods, we build our personalized parameterization approach PREP-SR as described below.

3.3.3 Construction of Feature Data Vector

To build PREP-SR model, we define a feature data vector which is composed of elements derived using moments of spatial randomness quadrat count and kernel estimation methods and *L-function* of random pattern analysis. In feature data vector, in addition to the parameters calculated using pattern analysis methods, we have included their correlations to take into account a possible dependency among these parameters. For a pathology slide ℓ_x, m dimensional feature data vector $\mathfrak{R}_{\ell_x}$ can be expressed as:

$$\mathfrak{R}_{\ell_x} = < c, \mathfrak{r}_1, \cdots, \mathfrak{r}_q, \zeta_1, \cdots, \zeta_\kappa > \quad (6)$$

where c is the parameter of offset constant, q is the number of spatial randomness features, κ is the number of correlation parameters, and $\mathfrak{r}_i$ and ζ_j stand for i^{th} spatial randomness and j^{th} correlation parameters, respectively. With the definition of set R for q number of features that are initially derived from spatial randomness methods

and each denoted as $\mathfrak{r}$, correlation parameters can be computed as:

$$\zeta_j = \prod_{\mathfrak{r}_i \in S_j} \mathfrak{r}_i \tag{7}$$

where the j^{th} correlated feature set S_j can be stated as:

$$S_j = \{\mathfrak{r}_i \in R: \quad \mathfrak{C}(\mathfrak{r}_i, \mathfrak{r}_k) \geq \tau, \ \forall \mathfrak{r}_k \in S_j \backslash \mathfrak{r}_i\} \tag{8}$$

where $\mathfrak{C}$ is the correlation function and τ is the threshold value to decide if there is a correlation among several features. In our study, we obtained the most useful results when set S_j was constructed with correlated pairs (i.e., $|S_j| = 2$) and τ was selected as 0.9. Note that parameter $\mathfrak{r}_i$ is a function formulated using either quadrat dimensions or kernel radius of first order properties, or it is the *L-function* of second order properties of spatial pattern analysis.

3.3.4 Feature Accordance Matrix $\mathfrak{A}_{SR}$

The impact of features extracted from H&E slides are identified by $(\mu \times m)$ dimensional feature accordance matrix $\mathfrak{A}_{SR} = [a_{ij}]$ which is formed for m number of feature vector elements and μ number of growth model parameters. In PREP-SR method, using coefficient of matrix $\mathfrak{A}_{SR}$, the vector of tumor growth parameters for a pathology slide ℓ_x is calculated as

$$K_{\ell_x} = \mathfrak{A}_{SR} \cdot \mathfrak{R}_{\ell_x} \tag{9}$$

where μ number of tumor growth parameters are denoted by vector $K_{\ell_x} =< k_1, k_2, \cdots, k_\mu >$ for slide ℓ_x.

To compute the elements of coefficient matrix $\mathfrak{A}_{SR}$, we have developed our DE-based algorithm using artificial intelligence techniques from our earlier research [32, 33, 34]. The process of computing the coefficient matrix $\mathfrak{A}_{SR}$ and tumor growth parameters is illustrated with Algorithm 1, where the input data is provided from tumor volume measurements of patient derived mice xenograft models and H&E slides collected at the end of the experiment.

In Algorithm 1, the candidate solutions are named as vectors, which are composed of feature index parameters and candidate coefficients of row vector $\mathfrak{A}^{\phi}_{SR}$. In this process, we define feature index parameters as part of each candidate vector to choose the dimension and radius for quadrat count, kernel estimation, and *L-function* of nearest neighbor (shown at line 4). The other elements of each candidate vector are defined as the coefficients of $\mathfrak{A}^{\phi}_{SR}$ row vector to weigh the computed features, as shown at line 5. Once the dimension and radius sizes and $\mathfrak{A}^{\phi}_{SR}$ row vector are calculated using the candidate, we compute feature data vector and weighted least square difference for each slide sample as shown in lines 6 to 10. Here, to decrease the variance of error rates among slide samples, we utilize a least square error calculation weighted inversely with the magnitude of the growth parameter. For a sample slide ℓ_x it is formulated as:

$$\varepsilon(\ell_{\phi,x}) = \left(\frac{k_{\phi,x} - \tilde{k}_{\phi,x}}{k_{\phi,x}} \right)^2 \tag{10}$$

where $k_{\phi,x}$ and $\tilde{k}_{\phi,x}$ are the measured and estimated values of parameter k_ϕ, respectively, for slide ℓ_x. Using error rates of each slide, we compute the fitness of candidate $\vec{\mathfrak{h}}^{g}_{i}$ for i^{th} individual in g^{th} generation as:

$$f(\vec{\mathfrak{h}}^{g}_{i}) = \sum_{\ell_x \in S} \varepsilon^{g}_{i}(\ell_x) \tag{11}$$

We evolve the candidate vector population to the next generation as illustrated at lines 13-16 of Algorithm 1. At line 13, the best population is determined using the fitness function, which approaches to a desirable solution due to exploitation feature of DE, where a more thorough search is performed over a local region. Exploration feature of DE, where a more pronounced random search over the global solution space takes place, is implemented at line 15. We repeat the procedure generation by generation to find out the fittest solution for $\mathfrak{A}^{\phi}_{SR}$ row vector (shown with lines 18-19). The fittest candidate $\vec{\mathfrak{h}}$ is calculated from solution space H as:

$$\vec{\mathfrak{h}} = \underset{\vec{\mathfrak{h}}^{g}_{i} \in H}{\operatorname{argmin}} \ f(\vec{\mathfrak{h}}^{g}_{i}) \tag{12}$$

At the end of Algorithm 1, the row vectors computed for each growth parameter $k_\phi \in K$ constitute the coefficient matrix $\mathfrak{A}_{SR}$ as shown at line 21.

Algorithm 1: Computation of matrix $\mathfrak{A}_{SR}$ and tumor growth parameters for H&E slide samples

Data: H&E slides and tumor growth parameters $k_\phi \in K$ from experimental volume measurements for n slides

1 **foreach** *tumor growth parameter* $k_\phi \in K$ **do**
2 **foreach** *candidate vector population* H^g *of generation* g **do**
3 **foreach** *candidate* $\vec{\mathfrak{h}}^{g}_{i} \in H^g$ **do**
4 define feature index parameters of $\vec{\mathfrak{h}}^{g}_{i}$;
5 define candidate A^{g}_{i} row vector of $\vec{\mathfrak{h}}^{g}_{i}$;
6 **foreach** *slide* $\ell_x \in S$ **do**
7 compute feature data vector $\mathfrak{R}_{\ell_x}$;
8 compute estimated $\tilde{k}_{\phi,x}$ value using Eq. (9);
9 calculate error $\varepsilon^{g}_{i}(\ell_{\phi,x})$ using Eq. (10);
10 **end**
11 calculate fitness using Eq. (11);
12 **end**
13 calculate fittest candidate vector population H^{g}_{best} ;
14 **if** $g < |G|$ **then**
15 evolve candidate population H^{g+1} from H^{g}_{best} ;
16 **end**
17 **end**
18 calculate fittest candidate vector $\vec{\mathfrak{h}}$ with Eq. (12) ;
19 define the $\mathfrak{A}^{\phi}_{SR}$ row vector of the fittest chromosome;
20 **end**
21 construct $\mathfrak{A}_{SR}$ matrix using $\mathfrak{A}^{\phi}_{SR}$ row vectors;

S: set of slides, $|G|$: number of generations, H: solution space

3.3.5 Computation of tumor growth parameters

We utilize spatial randomness features to compute tumor growth parameters for a given set of H&E slides. In our personalized parameterization model PREP-SR, array of growth parameters $\widetilde{K}^{\phi}$ for n number of H&E slides is denoted as:

$$\widetilde{K}^{\phi} = (\tilde{k}_{\phi,\ell_1}, \ \tilde{k}_{\phi,\ell_2}, \ \cdots, \ \tilde{k}_{\phi,\ell_n}) \tag{13}$$

where $\tilde{k}_{\phi,\ell_x}$ is the value of the ϕ^{th} parameter for slide ℓ_x. The equation set to compute $\mathfrak{A}^{\phi}_{SR}$ row vector of matrix $\mathfrak{A}_{SR}$ corresponding to ϕ^{th} growth parameter of all slide samples can be formulated as:

$$\mathfrak{A}^{\phi}_{SR} = \widetilde{K}^{\phi} \cdot \begin{bmatrix} C \\ \mathfrak{R}_R \end{bmatrix}^{-1} \tag{14}$$

where C is a row vector whose elements consist of offset constant c for n slides, and $\mathfrak{R}_R$ is the reduced form of matrix $\mathfrak{R}$.

Using Eq. 14, the computation of tumor growth parameter k_ϕ for sample slide ℓ_x can be stated as:

$$k_\phi = c \cdot a_{\phi,0} + \sum_{s=1}^{q} a_{\phi,s} \cdot r_s + \sum_{t=1}^{\kappa} a_{\phi,(t+q)} \cdot \zeta_t \qquad (15)$$

where $a_{\phi,j}$ is the j^{th} coefficient of the row vector $\mathfrak{A}_{SR}^{\phi}$. Note that the number of correlation parameters included in the feature data vector is restricted by a constant ρ_c to obtain an over-determined system of equations to be used in PREP-SR. For this reason, in Algorithm 1, we disregard the candidates which result in more correlation parameters than ρ_c. In the study presented in this paper, considering the values of the total number of spatial randomness parameters and the total number of slides, we set $\rho_c = 4$.

3.3.6 *Cross validation of* PREP-SR

In PREP-SR, tumor volume measurements and corresponding H&E stained slides are used to compute the tumor growth parameters. We apply a leave-one-out-cross-validation (LOOCV) [11] based technique as a performance metric to evaluate the outputs generated by PREP-SR. Implementation of this procedure is presented in Algorithm 2.

Algorithm 2: Cross validation algorithm for $\mathfrak{A}_{SR}$ matrix

Data: H&E slides and tumor growth parameters $k_\phi \in K$ from experimental volume measurements for n slides

1 **foreach** *slide* $\ell_u \in S$ **do**
2 remove slide ℓ_u from set S : $S_u = S \backslash \ell_u$;
3 remove parameters of vector K_{ℓ_u} from set K: $K_u = K \backslash K_{\ell_u}$;
4 compute $\mathfrak{A}_{SR}(\ell_u)$ matrix using sets S_u and K_u in Algorithm 1;
5 compute feature data vector $\mathfrak{R}_{\ell_u}$ using fittest candidate $\mathfrak{h}$ from Algorithm 1;
6 compute vector K_{ℓ_u} using Eq. (9);
7 calculate error rate in tumor volume with $\varepsilon_{\mathrm{V}}(\ell_u) = \sum_{t_i \in \mathrm{T}_u} |\mathrm{V}_{\phi,u}(t_i) - \mathrm{V}_{\phi,u,comp}(t_i)| \;/\; \mathrm{V}_{\phi,u}(t_i)$
8 **end**
9 calculate cross validation error rate $\varepsilon_{\mathrm{CV}} = \frac{1}{|S|} \cdot \sum_{\ell_u \in S} \varepsilon_{\mathrm{V}}(\ell_u)$

In Algorithm 2, at lines 2-3, we define a training set from n slides to construct feature accordance matrix $\mathfrak{A}_{SR}(\ell_u)$, which corresponds to slide ℓ_u. Once PREP-SR computes $\mathfrak{A}_{SR}(\ell_u)$ matrix using Algorithm 1, we compute feature data vector $\mathfrak{R}_{\ell_u}$ in lines 4-5 using the fittest candidate $\vec{\mathfrak{h}}$. In line 6, the growth parameter vector K_{ℓ_u} is formed using Eq. (9). In line 7, with the computed parameters for each slide ℓ_u, tumor volume can be calculated for each time point $t_i \in \mathrm{T}_u$ where T_u is defined as the set of time points at which preclinical measurements of slide ℓ_u are recorded.

The error rates of tumor volume values from PREP-SR model are calculated as the difference between these computed values and the corresponding measurements recorded during preclinical study. The percentage error rate for slide $\ell_u \in S$ and the mean error rate for all slides of set S are shown at lines 7 and 9 of Algorithm 2, respectively.

4. ANALYTICAL RESULTS

In this study, we compute tumor growth parameter λ_1 of the exponential-linear model with preclinical measurements taken after volume of tumor mass reached 400 mm^3. With the construction of feature accordance matrix $\mathfrak{A}_{SR}$, the LOOCV based methodology in Algorithm 2 is used to evaluate the error rate. In Fig. 3, we present

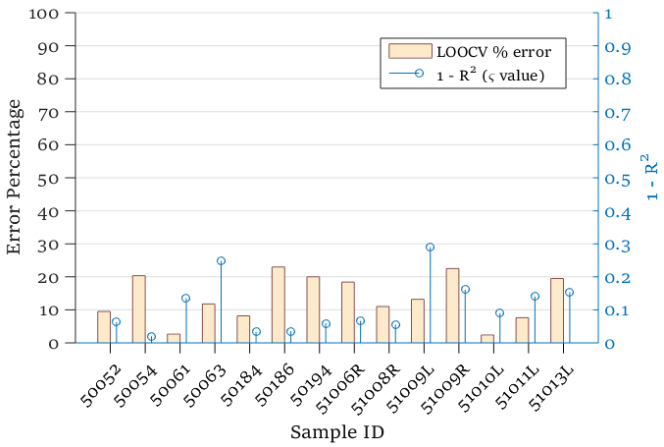


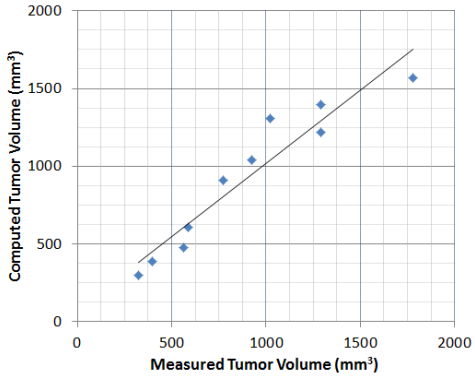
Figure 3: The LOOCV percentage error rate of H&E slide samples and significance value for measured and computed tumor volumes for each slide

the error rates with bars for slide samples, where the percentage of error rate is scaled with the vertical axis on the left side. Based on PREP-SR results, we observe an average LOOCV error rate of 13.58% for 14 sample H&E slides.

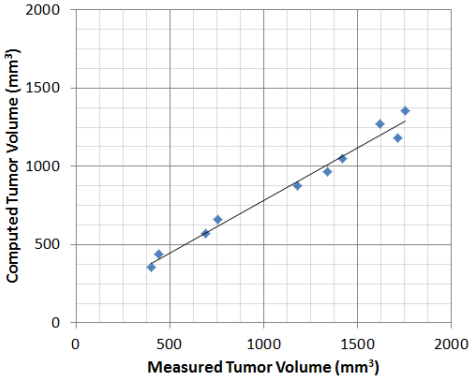
In addition, we can analyze the significance of the error rate for PREP-SR model by computing the coefficient of determination (denoted as $R^2 \in [0,1]$), which is calculated as the square of the correlation between two sets of values. The inverse of significance level ς is formulated as $\varsigma = 1 - R^2$, where R^2 is the coefficient of determination for the computed and measured volumes of tumor mass at multiple time points during the experiment. The inverse significance level will be low when the correlation is high, implying that the measurements taken during preclinical experiment are consistent with exponential-linear growth behavior. In Fig. 3, inverse significance values for 14 sample slides are shown with the vertical axis on the right hand side.

To provide a closer look at our analytical results, let us examine four different samples with various significance levels and error rates chosen from Fig. 3. The scatter graphs shown in Figs. 4a - 4d are for the H&E slide samples of 51010-L, 50054, 51013-L and 50052, respectively. An ideal result can be defined as a case where the computed and the measured values are the same (i.e., slope of the trendline is 1), and the data points are all on the trendline (i.e.. $R^2 = 1$). The four samples are selected such that the results show a low error rate with a high significance for Sample 51010-L (Fig. 4a), high error rate with high significance for 50054 (Fig. 4b), high error rate with low significance for 51013-L (Fig. 4c), and an average error rate with an average significance for 50052 (Fig. 4d).

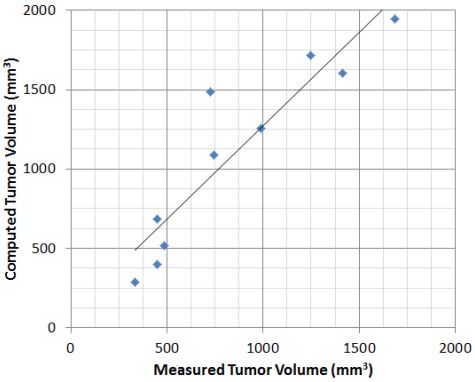
In Figs. 4a - 4d, the difference between the slope of the trendline and an ideal slope of 45° rising line is proportional to the average error rate from Fig. 3. The scatter graph of correlation is depicted for Sample 51010-L in Fig. 4a, for which we obtained an error rate of 2.5% with an inverse significance value of $\varsigma = 0.09$. The deviation of trendline's slope (calculated as 0.95) from 45° rising line is small, which is consistent with the small error rate for this sample. In Fig. 4b, for the correlation scatter graph of sample 50054, we observe that the error rate is relatively larger (calculated as 20.35%) with a high significance ($\varsigma = 0.02$ in Fig 3). In Fig. 4c, for Sample 51013-L, we observe a similar error rate (calculated as 19.5%) but a lower significance ($\varsigma = 0.15$) than for Sample 50054. As can be seen in Fig. 4c, the computed volumes are larger than the preclini-



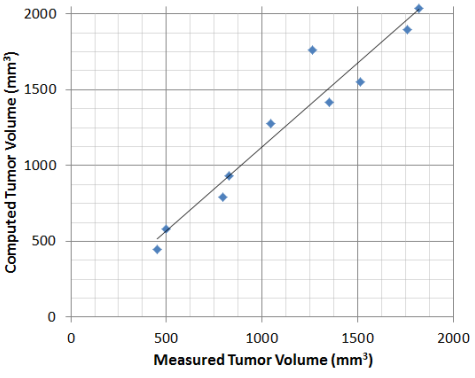
(a) Sample Slide 51010-L



(b) Sample Slide 50054



(c) Sample Slide 51013-L



(d) Sample Slide 50052

Figure 4: The Scatter Plot of correlation between measured and computed tumor volumes and the trendlines for H&E sample slides

cal measurements for all time points, which results in a slope value larger than 1. Fig. 4d illustrates a scatter plot of Sample 50052, for which an approximately average result (9.53%) was obtained with an average significance ($\varsigma = 0.06$) in our study.

Table 2: Comparison of LOOCV % error rate, inverse significance value $\varsigma = (1 - R^2)$, and slope of trendline for the selected H&E slides of 51010-L, 50054, 51013-L and 50052

Results from Figs. 3 & 4				
Metric Type	51010-L	50054	51013-L	50052
% error rate	2.50	20.35	19.50	9.53
$1 - R^2$	0.09	0.02	0.15	0.06
slope of trendline	0.95	0.67	1.18	1.11

Values of percentage error rate, inverse significance and the slope of trendline are given in Table 2 for the sample slides of Fig. 4. We see that the error rate increases as the difference between the slopes of trendline and 45° rising line gets larger (as can be seen from Samples 51010-L and 50052, which have close significance values). Another observation is that the dependency between error rate and the slope difference is influenced by the significance value (as can be seen from Samples 50054 and 51013-L, which have close error rates).

5. CONCLUDING REMARKS

In this paper, we present PREP-SR method to compute personalized tumor growth parameters based on the digitized images of H&E slides equipped with kidney cancer tumor tissues from patient derived xenograft models. To examine the hypothesis that the nuclei positions possess information of tumor growth behavior specific to slides, we extract features using spatial pattern analysis techniques of quadrat counts, kernel estimation and nearest neighbor functions. Using spatial randomness features as inputs to PREP-SR, we compute tumor growth parameters for exponential-linear model. The DE-based algorithm in PREP-SR defines a candidate vector consisting of size selection indices for spatial evaluation and weight coefficients for spatial features and their correlations. The fittest candidate vector is employed to compute the slide-specific parameters for new slide samples. With LOOCV method, for a set of H&E slides obtained from xenograft models implanted with kidney cancer tumor tissues, PREP-SR generates personalized model parameters with an average error rate of 13.58%. Also, the level of correlation between preclinical measurements and the computed results indicates that the measurements taken during preclinical experiment are consistent with exponential-linear growth behavior.

These promising results encourage development of artificial intelligence based methods utilizing spatial features to compute personalized tumor growth parameters for mathematical models. These models based on patient specific information may play an important role in supporting clinical systems. We plan to extend our research to include computation of tumor shrinkage parameters for kidney cancer patients.

6. REFERENCES

[1] M. Simeoni, P. Magni, C. Cammia, et al. Predictive pharmacokinetic pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer Research*, 64:1094–1101, 2004.

[2] Emir Ganic, Stephen Gundry, Jianmin Zou, and M. Umit Uyar. Evaluation of anti-cancer therapy using in silico

analysis of treatments for HER2+ breast cancer. In *IEEE Intl. Symp. on Medical Measurements and Applications (MeMeA 2014)*, pages 211–216, Lisbon, Portugal, June, 2014.

[3] J. Harrold and R.Parker. Clinically relevant cancer chemotherapy dose scheduling via mixed-integer optimization. *Computers & Chemical Engineering*, 33(12):2042–2054, 2009.

[4] Gilbert Koch, Antje Walz, Gezim Lahu, et al. Modeling of tumor growth and anticancer effects of combination therapy. *Journal of Pharmacokinetcs and Pharmacodynamics*, 36:179–197, 2009.

[5] Metin N Gurcan, Laura E Boucheron, Ali Can, Anant Madabhushi, Nasir M Rajpoot, and Bulent Yener. Histopathological image analysis: A review. *Biomedical Engineering, IEEE Reviews in*, 2:147–171, 2009.

[6] Aydin Saribudak, Yiyu Dong, James Hsieh, and M Umit Uyar. Modeling tumor growth for kidney cancer based on nuclei clusters of pathology slides. *International Journal of Engineering and Technology*, 8(5), 2016.

[7] Brian D Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 172–212, 1977.

[8] Anthony C Gatrell, Trevor C Bailey, Peter J Diggle, and Barry S Rowlingson. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British geographers*, pages 256–274, 1996.

[9] P. Haase. Can isotropy vs. anisotropy in the spatial association of plant species reveal physical vs. biotic facilitation? *Jour. of Vegetation Science*, 12(1):127–136, 2001.

[10] Kenneth Price, Rainer M Storn, and Jouni A Lampinen. *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, 2006.

[11] Sylvain Arlot, Alain Celisse, et al. A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79, 2010.

[12] Katarzyna A Rejniak and Alexander RA Anderson. Hybrid models of tumor growth. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 3(1):115–125, 2011.

[13] L. Bueno, D. P de Alwis, C. Pitou, J. Yingling, M. Lahn, S. Glatt, and I. F Trocóniz. Semi-mechanistic modelling of the tumour growth inhibitory effects of ly2157299, a new type i receptor tgf-β kinase antagonist, in mice. *European journal of cancer*, 44(1):142–150, 2008.

[14] N. Orlov, L. Shamir, T. Macura, J. Johnston, D Mark Eckley, and I. G Goldberg. Wnd-charm: Multi-purpose image classification using compound image transforms. *Pattern recognition letters*, 29(11):1684–1693, 2008.

[15] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In *Proc. 5th IEEE ISBI*, pages 496–499, 2008.

[16] P. J. Diggle. On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics*, pages 87–101, 1979.

[17] D. U. Pfeiffer. Issues related to handling of spatial data. *Massey University, Palmerston North*, pages 309–329, 1996.

[18] G. L. W. Perry, B. P. Miller, and N. J Enright. A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. *Plant ecology*, 187(1):59–82, 2006.

[19] Adrian Baddeley and Rolf Turner. Modelling spatial point patterns in r. In *Case studies in spatial point process modeling*, pages 23–74. Springer, 2006.

[20] Aydin Saribudak, Emir Ganic, Jianmin Zou, Stephen Gundry, and M. Umit Uyar. Toward genomic based personalized mathematical models for breast cancer tumor growth. In *IEEE 14th Int'l. Conf. on BioInformatics and BioEngineering (BIBE)*, pages 115–119, Nov. 2014.

[21] Aydin Saribudak, Stephen Gundry, Jianmin Zou, and M. Umit Uyar. Genomic based personalized chemotherapy analysis to support decision systems for breast cancer. In *Medical Measurements and Applications (MeMeA), 2015 IEEE Intl. Symp. on*, pages 495–500. IEEE, 2015.

[22] R. Edgar, M. Domrachev, and A. E Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic acids research*, 30(1):207–210, 2002.

[23] NBIA - National Biomedical Imaging Archive. *National Institutes of Health, https://imaging.nci.nih.gov/ncia*.

[24] John W Park, Keelung Hong, Dmitri B Kirpotin, Gail Colbern, Refaat Shalaby, Jose Baselga, Yvonne Shao, Ulrik B Nielsen, James D Marks, Dan Moore, et al. Anti-her2 immunoliposomes enhanced efficacy attributable to targeted delivery. *Clinical Cancer Research*, 8(4):1172–1181, 2002.

[25] S. Man, G. Bocci, G. Francia, S. K Green, S. Jothy, D. Hanahan, P. Bohlen, D. J Hicklin, G. Bergers, and R. S Kerbel. Antitumor effects in mice of low-dose (metronomic) cyclophosphamide administered continuously through the drinking water. *Cancer Research*, 62(10):2731–2735, 2002.

[26] Chuanren Liu, Tianming Hu, Yong Ge, and Hui Xiong. Which distance metric is right: An evolutionary k-means view. In *SDM*, pages 907–918. SIAM, 2012.

[27] Maurizio Filippone, Francesco Camastra, Francesco Masulli, and Stefano Rovetta. A survey of kernel and spectral methods for clustering. *Pattern recognition*, 41(1):176–190, 2008.

[28] A. K. Jain, M. N. Murty, and P. J Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323, 1999.

[29] Tony E Smith. A scale-sensitive test of attraction and repulsion between spatial point patterns. *Geographical Analysis*, 36(4):315–331, 2004.

[30] Brian D Ripley. The second-order analysis of stationary point processes. *Journal of applied probability*, pages 255–266, 1976.

[31] L. Norton. Conceptual and practical implications of breast tissue geometry: toward a more effective, less toxic therapy. *The Oncologist*, 10(6):370–381, 2005.

[32] J. Kusyk, C.S. Sahin, M.U. Uyar, E. Urrea, and S. Gundry. Self organization of nodes in mobile ad hoc networks using evolutionary games and genetic algorithms. *J. of Advanced Research*, 2(3):253 – 264, 2011.

[33] S. Gundry, J. Zou, E. Urrea, C. S. Sahin, J. Kusyk, and M. U. Uyar. Analysis of emergent behavior for ga-based topology control mechanism for self-spreading nodes in MANETs. In *Advances in Intelligent Modelling and Simulation*, volume 422 of *Studies in Computational Intelligence*, pages 155–183. Springer Berlin Heidelberg, 2012.

[34] C.S. Sahin, M. U. Uyar, S. Gundry, and E. Urrea. Self organization for area coverage maximization and energy conservation in mobile ad hoc networks. In *Transactions on Computational Science XV*, pages 49–73. Springer, 2012.