Forecasting new development of tumor areas using spatial and temporal distribution profiles of hemoglobin saturation in a mouse model

Miguel R. Ossandon
Dhananjay S. Phatak
Brian S. Sorg
Konstantinos Kalpakis

Forecasting new development of tumor areas using spatial and temporal distribution profiles of hemoglobin saturation in a mouse model

Miguel R. Ossandon,^{a,b,*} Dhananjay S. Phatak,^a Brian S. Sorg,^b and Konstantinos Kalpakis^a

^aUniversity of Maryland Baltimore County, Department of Computer Science and Electrical Engineering, 1000 Hilltop Circle, Baltimore, Maryland 21250

^bNational Cancer Institute, 9609 Medical Center Drive, Rockville, Maryland 20850

Abstract. Features of the tumor microenvironment (TME), such as hemoglobin saturation (HbSat), can provide valuable information on early development and progression of tumors. HbSat correlates with high metabolism and precedes the formation of angiogenic tumors; therefore, changes in HbSat profile can be used as a biomarker for early cancer detection. In this project, we develop a methodology to evaluate HbSat for forecasting early tumor development in a mouse model. We built a delta (δ) cumulative feature that includes spatial and temporal distribution of HbSat for classifying tumor/normal areas. Using a two-class (normal and tumor) logistic regression, the δ feature successfully forecasts tumor areas in two window chamber mice (AUC = 0.90 and 0.85). To assess the performance of the logistic regression-based classifier utilizing the δ feature of each region, we conduct a 10-fold cross-validation analysis (AUC of the ROC = 0.87). These results show that the TME features based on HbSat can be used to evaluate tumor progression and forecast new occurrences of tumor areas. © 2014 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.1.1.014503](https://doi.org/10.1117/1.JMI.1.1.014503)]

Keywords: tumor angiogenesis; image analysis; tumor microenvironment; hemoglobin saturation.

Paper 13013RR received Nov. 5, 2013; revised manuscript received Apr. 20, 2014; accepted for publication May 19, 2014; published online Jun. 20, 2014.

1 Introduction

Tumor progression and development can alter local metabolism, producing physiological changes in the tumor microenvironment (TME). Features associated with TME, such as development of new vessels, dilatation, deformation or destruction of existing vessels, oxygenation and hemoglobin saturation (HbSat), can provide valuable information on the progression of the tumor.^{1,2} For example, high metabolism, which is the characteristic of many cancers, produces early changes in levels of oxygenation around the tumor resulting in local changes of HbSat in tumor microvessels followed by a decrease in oxygenation in the central part of the tumor.³ Several studies have shown that the HbSat is useful in distinguishing tumor areas from both normal fibro-glandular and adipose tissues.^{4,5} Other studies have shown that in spite of the intense angiogenesis, tumors can still be underserved with oxygen due to hyper metabolism of the cancer.^{4–6} As the tumor progresses, the central low-oxygenated internal part of the tumor grows and microvessels around the margin of the tumor have a significant increase in HbSat.³ This spatial distribution of HbSat is iterated because the tumor grows creating an HbSat profile from tumor proliferation to necrosis (areas with low HbSat). These changes in oxygenation and spatiotemporal distribution of the HbSat can provide insight in overall tumor progression^{1,7} and can be helpful in early tumor detection or the evaluation of response to treatment.

Figure 1 shows a picture of a tumor developed from 4T1 mouse mammary carcinoma cells injected in a dorsal skin-fold window chamber in an athymic nude mouse. The picture was taken 4 days after the implantation of 5000 tumor cells. The tumor cells were transfected with reporter genes of red fluorescence protein (RFP) for identification of the tumor location. The RFP vector was engineered by splicing the RFP coding sequence cloned from pDsRed2-N1 into pLPCX (BD Biosciences, San Jose, California) by EcoRI-NotI digestions as described in Ref. 8.

In Fig. 1(b), we can appreciate the high levels of saturation in the margin of the tumor compared with lower levels of saturation in the central part of the tumor (light blue). Outside the margin of the tumor, we can see high levels of HbSat as well. These areas of high HbSat outside the margin of the tumor indicate increased metabolism consistent with future tumor growth and show that the HbSat preceded the formation of the tumor. Analysis of temporal and spatial distributions of HbSat profile over the early development of the tumor can be useful for early cancer detection.

2 Aims and Contributions

The aim of this project is to develop a methodology based on image registration and segmentation for the analysis of HbSat for forecasting tumor development in a dorsal skin-fold window chamber. Toward this end, we use the spatial and temporal distribution of HbSat, though other features of TME can be analyzed in the same way.

*Address all correspondence to: Miguel R. Ossandon, E-mail: ossandon@mail.nih.gov

Fig. 1 (a) Tumor grown in dorsal skin-fold window chamber 4 days after tumor cell implantation. (b) The margin of the tumor is represented in black. The pixel values were replaced by HbSat values and displayed as a color map. Note the high hemoglobin saturation values (orange-red) in areas surrounding the tumor margin.

We propose a new delta feature (δ) computed from the HbSat data over time, and showed that it is effective at forecasting future tumor occurrence in mouse models. In particular, a logistic regression-based classifier using our proposed new feature achieves $AUC = 0.87$ in our cross-validation experimental study. This is potentially useful for early cancer detection.

3 Methodology

In this project, we work with two sets of images for each day during 5 days, gray-scale images [Fig. 2(a)] and data images [Fig. 2(b)]. The gray-scale image is a transmitted light image from which structural characteristics (e.g., vascularity) of tissue inside the window chamber can be evaluated. A data image is a pixel-wise mapping of HbSat values of blood vessels in the image.

The image datasets were obtained prior to this project by one of the authors in the laboratory of Mark W. Dewhirst (Duke University, Durham, NC). The procedure used to obtain the image data is described in Ref. 7.

Spatial and temporal distribution of HbSat was determined by segmenting the images in windows of 20×20 pixels covering normal and tumor areas. The pixel resolution was about

$3.5 \mu\text{m}$ with a pixel area of $12.25 \mu\text{m}^2$ (the pixels were square). The 20×20 windows' size was selected after testing different window sizes (20, 50, and 100). The 20×20 windows' size was determined to be the most cost effective for this project. Bigger windows are easy to map and require less computational power, but smaller windows provide more accurate representation of the HbSat in the targeted region. Corresponding window areas over several days during the development of the tumor were compared to provide an HbSat profile. This information was used later to predict a tumor development in areas with no occurrence of tumor cells as yet.

The areas of newly developed tumors were determined by subtracting from the tumor areas of a particular day the tumor areas of previous days as follows:

$$T_j^{\text{new}} = T_j^{\text{all}} - T_{(1:i)}^{\text{all}}; \quad i < j, \quad (1)$$

T_j^{new} = New tumor are as developed at day j ; T_j^{all} = All tumor are as developed at day j ; $T_{(1:i)}^{\text{all}}$ = All tumor are as from day 1 to i .

The vector " T " in Eq. (1) represents a one-dimensional (1-D) vector containing the labels of the tumor areas at particular day, for example:

Fig. 2 Gray-scale image (a) and data image (b). The picture was taken 1 day after the implantation of the tumor cells. The transmitted light image (a) is represented as a gray-scale image. The data image (b) contains the hemoglobin saturation values.

$T5^{all}$ represents all tumor areas at day 5: $T5^{all}$
 $= \langle 39 \ 41 \ 55 \ 63 \ 82 \ 99 \ 106 \ 109 \ 201 \rangle$

$T4^{all}$ represents all tumor areas at day 4: $T4^{all}$
 $= \langle 39 \ 55 \ 63 \ 99 \ 109 \ 201 \rangle$

$T5^{new} = T5^{all} - T4^{all}$
 $\Rightarrow \begin{array}{c} \langle 39 \ 41 \ 55 \ 63 \ 82 \ 99 \ 106 \ 109 \ 201 \rangle \\ - \langle 39 \ 55 \ 63 \ 99 \ 109 \ 201 \rangle \\ \hline = \langle 41 \ 82 \ 106 \rangle \end{array}$

$T5^{new} = \langle 41 \ 82 \ 106 \rangle$

Areas 41, 82, and 106 are new areas of tumor developed at day 5 (not present before day 5).

Areas 39, 55, 63, 99, 109, and 201 are areas of tumor at day 5 that were developed before day 5.

4 Registration

Since the structure of tissue and vascularity is not evident in the data images, the registration is performed in the gray-scale images and then projected to the data images for the purpose of HbSat evaluation. For the registration steps, we developed a graphical user interface “regGUI” (Fig. 3) based on the MATLAB® functions “cp2tform,” “imtransform,” and “imshowpair.”

- The “cp2tform” function infers a spatial transformation from control point pairs (pixel coordinates) and returns this spatial transformation in a structure “t_concord” (see below).

$t_concord = cp2tform(xyinput_out, xybase_out, 'nonreflective similarity');$

Fig. 3 regGUI. The moving image is displayed in red.

*xyinput_out: control points from the unregistered gray-scale image

**xybase_out: control points from the base gray-scale image

The control points (xyinput_out, xybase_out) were visually selected by translation and rotation of the unregistered image over the base image using the “regGUI” tool. The control points are pixel coordinates of similar features in both images.

- The MATLAB function “imtransform” transforms an image according to 2-D spatial transformation defined by the “t_concord” structure and returns a registered image (GSreg) (Fig. 4).

$GSreg = imtransform(input, t_concord, 'XData', [1 Width], 'YData', [1 Height]);$

Using the same control point pairs (xyinput_out, xybase_out), we can also reverse the registration process switching their positions in “cp2tform” producing the structure “t_concordReversed” that carries the spatial transformation information to reverse the registration. We call this process deregistration.

$t_concordReversed = cp2tform(xybase_out, xyinput_out, 'nonreflective similarity');$

The structures “t_concord” and “t_concordReversed” are used in subsequent steps for mapping of the segmented areas (20×20 pixels) of the base data image with the unregistered data image.

4.1 Segmentation and Mapping

The segmentation of the images is based on a mesh of 20×20 pixel windows with the same dimensions of the base image. One copy of the mesh is deregistered using function “imtransform” with “t_concordReversed” as a parameter, as shown in Fig. 5.

$GSreg = imtransform(mesh, t_concordReversed, 'XData', [1 Width], 'YData', [1 Height]);$

Another copy of the mesh is overlaid in the base image (Fig. 6, upper half). The deregistered mesh from the previous step is overlaid in the unregistered image; in this case, the data image is for day 2 (Fig. 6, lower half).

Subsequently, the base data and the unregistered data images (both with the overlaid mesh) are segmented/labeled based on the overlaid mesh (Fig. 7) using MATLAB functions “bwconncomp” and “labelmatrix.”

$conectedComp = bwconncomp(image_Data + Mesh);$
 $labelImage = labelmatrix(conectedComp);$

Fig. 4 Registration of the image at day 2.

Fig. 5 A copy of the mesh is deregistered using the spatial transformation provided by “t_concordReversed”

Fig. 6 The deregistered mesh is overlaid on the unregistered data image for day 2 (lower half), the original mesh is overlaid over the data image for day 1 (upper half).

Using the original “t_concord” structure as the parameter in function “imtransform,” the labeled image day 2 is registered (Fig. 8) to match with the labeled image day 1.

Labeledreg=imtransform(Labeled-image-day-2, t_concord, 'XData', [1 Width], 'YData', [1 Height]);

Since the registration was performed using the original “t_concord” structure, the windows of the mesh can be mapped in both images (days 1 and 2) using their center of mass (Fig. 9) (MATLAB function “regionprops” with the centroid property). The corresponding regions for each day are stored in a matrix (Regions

Fig. 7 The image for days 1 and 2 is segmented based upon the overlaid mesh.

Fig. 8 The labeled image day 2 is registered using the original “t_concord” structure.

Fig. 9 Using the center of mass, the windows in the labeled base data image for day 1 on the left and labeled image for day 2 on the right can be mapped. For example, window 47 in day 1 corresponds to window 42 in day 2.

Table 1 Sample of the mapped registered regions of the images. For example, region 1688 in day 1 is equivalent to region 1636 on day 2 and region 1442 on day 3, and so on.

Region Day 1	Region Day 2	Region Day 3	Region Day 4	Region Day 5
1688	1636	1442	1306	1306
1689	1637	1443	1307	1307
1690	1638	1444	1308	1308
1691	1639	1445	1309	1309
1692	1640	1446	1310	1310
1693	1641	1447	1311	1311
1694	1642	1448	1312	1312
1695	1643	1449	1313	1313
1696	1644	1450	1314	1314
1697	1645	1451	1315	1315
1698	1646	1452	1316	1316
1699	1647	1453	1317	1317

X Days), in which the regions correspond to the rows and the days correspond to the columns of the matrix as shown in Table 1.

5 Region Classification: Ground Truth

The 20 × 20 pixel regions in the images were classified as normal (0), tumor (1), or unknown (3) for each day (Fig. 10); the classes of the regions were derived.

Fig. 10 Selection of tumor (red) and normal (green) areas. Each area corresponds to 20 × 20 pixels.

Tumor areas were identified using fluorescence from tumor cells transfected with constitutively expressed reporter genes for RFP. A mask created by using the fluorescent image was overlapped on the image data to locate the tumor areas. Since changes in HbSat preceded the formation of the tumor, and normal areas were selected in a region of the window chamber far away from the original inception of the tumor cells (Fig. 10). All remaining areas were classified as unknown.

The classification of each region, e.g., normal, tumor, or unknown, was stored in a matrix similar to the mapped registered regions using day 1 as the reference for the region number (Table 2). In all the experiments described later in this paper, we discarded all regions labeled “Unknown.”

Table 2 Sample of classes assigned to the 20×20 pixels regions. For example, region reference 1692 is classified as unknown for days one and two, and classified as tumors for days three, four, and five.

Region	Day 1	Day 2	Day 3	Day 4	Day 5
1688	1	1	1	1	1
1689	3	1	1	1	1
1690	3	1	1	1	1
1691	3	1	1	1	1
1692	3	3	1	1	1
1693	3	3	3	1	1
1694	3	3	3	1	1
1695	3	3	3	1	1
1696	3	3	3	1	1
1697	3	3	3	3	1
1698	3	3	3	3	3
1699	3	3	3	3	3

5.1 Region Features

For each region/window r of 20×20 pixels in an image at day t , we fit, using maximum-likelihood estimation, a logistic distribution to the histogram of the region's pixel HbSat values. Recall that the density function of the logistic distribution is given by the following equation:

$$f(x|\mu, \sigma) = \frac{e^{-\frac{(x-\mu)}{\sigma}}}{\sigma \left[1 + e^{-\frac{(x-\mu)}{\sigma}} \right]^2}. \quad (2)$$

The estimated mean $\mu_t^{(r)}$ of this distribution is used as a feature of the region r at day t . We opted for the logistic distribution after a visual inspection of quality of fit of distributions to the histograms of the HbSat values of tumor and normal regions. The $\mu_t^{(r)}$ feature for each region r for every day t was stored in a matrix along with the corresponding matching region (Table 3).

5.2 Feature $\mu_t^{(r)}$

Since the $\mu_t^{(r)}$ value is a representation of the HbSat distribution of each 20×20 pixel window, it can be used to correlate HbSat with the tumor presence for each day. The feature $\mu_t^{(r)}$ was used to classify the regions as tumor or normal at day 4 [Fig. 11(a)] and at day 5 [Fig. 11(b)].

We classify regions as tumor or normal at day t using a two-class logistic regression classifier

$$\text{logit}(\text{odds-ratio of tumor in region } r \text{ at day } t) \sim 1 + \mu_t^{(r)},$$

which implies that the probability of a tumor in region r at day t is

$$\text{Prob}[\text{tumor in region } r \text{ at day } t] = \frac{e^{a_t \mu_t^{(r)} + b_t}}{1 + e^{a_t \mu_t^{(r)} + b_t}},$$

Table 3 Example of parameter μ for regions 1688 to 1699. The zero values represent regions with no valid reading of hemoglobin saturation (HbSat). As in Table 2, the region label from day one was used as reference.

Region	μ Day 1	μ Day 2	μ Day 3	μ Day 4	μ Day 5
1688	13.86044	44.69395	0	55.71582	31.24084
1689	17.21853	36.86726	43.76546	52.50144	0
1690	11.6666	27.13536	43.02122	0	0
1691	12.16857	32.53285	40.7397	0	7.275076
1692	20.19763	29.36109	41.28578	49.15093	0
1693	25.87345	0	41.26296	26.9634	30.13148
1694	15.66612	31.89307	42.31571	35.37616	0
1695	25.85196	36.09946	51.75413	54.93329	49.42748
1696	28.1438	50.29872	73.07992	58.82945	58.46795
1697	34.05655	29.76176	39.22287	58.22336	56.63576
1698	34.43235	23.98752	42.29988	54.93239	63.93473
1699	27.70158	25.65621	17.20532	0	72.11464

where the parameters a_t and b_t are estimated using the training data for day t . The logistic regression computes scores for each region, then we convert these scores to tumor/normal labels by comparing them to a threshold value. Then, we can assess the performance of this classifier by comparing the derived labels with the ground truth labels using receiver-operating-curve (ROC) analysis. The resulting ROC curves are shown in Fig. 11, together with the area under the curve (AUC), for classifying regions as tumor/normal in day 4 [Fig. 11(a)] and in day 5 [Fig. 11(b)].

We can see in Fig. 11 that using the parameter $\mu_t^{(r)}$ as a region's feature can discriminate very well between normal and tumor areas. These results were expected since there is previous research indicating that HbSat can distinguish malignant from normal tissue.⁴

6 Forecasting of Tumor Development in New Areas

As shown above, using $\mu_t^{(r)}$, it is possible to discriminate between tumor and normal areas in a particular day. However, since the increase of HbSat preceded the formation of tumor, it is interesting to know if $\mu_t^{(r)}$ can also predict for areas which lack tumor on a particular day, which new areas will become tumors in subsequent days due to tumor growth. For example: Is it possible to predict new tumor areas in days 4 or 5 using HbSat readings from day 3? To tackle this question, we use $\mu_t^{(r)}$ values from day 3 to classify areas as tumor in days 4 and 5. We compare our forecasting results with the ground truth classification provided by the fluorescence data. Tumor areas were selected using Eq. (1); normal areas were selected as previously described. As shown, we use logistic regression to classify regions as tumor/normal, and the results in ROC curves for the classification at days 4 and 5 using $\mu_3^{(r)}$ are given in Fig. 12.

Fig. 11 Correlation between μ values and tumor/normal areas on days 4(a) and 5(b) (x-axis: false-positive rate; y-axis: true-positive rate).

Fig. 12 Classification performance when using logistic regression on the $\mu_3^{(r)}$ values on day 3 to forecast tumor areas on days 4(a) and 5(b) (x-axis: false-positive rate; y-axis: true-positive rate).

Table 4 The areas selected in this table were classified as tumors in day 5 or/and day 4 according to the fluorescence maker (ground truth). In this example, all the $\mu_t^{(r)}$ values at day 3 (column 4) are under the threshold of 25 and therefore misclassified as normal.

Region	$\mu_{\text{Day 1}}$	$\mu_{\text{Day 2}}$	$\mu_{\text{Day 3}}$	$\mu_{\text{Day 4}}$	$\mu_{\text{Day 5}}$
1834	9.941389	20.37968	22.56743	70.88651	70.66735
1835	10.33063	16.63392	10.03903	41.08722	69.52692
1887	11.6746	20.2385	19.68554	54.97773	66.50639
1911	11.0895	16.34739	19.80297	57.02607	64.28368
1962	17.26394	22.93018	24.83022	56.97746	49.11736
2015	17.26017	19.00166	23.63816	58.27485	0
2046	24.22178	29.7639	21.45054	53.06991	0
2074	6.825127	8.281687	23.69352	60.88362	80.29969
2075	10.69403	13.19983	15.34123	62.56251	78.66432
2076	15.10585	11.10041	15.90866	64.54861	92.58891
2127	5.905925	7.681783	24.64143	57.3209	0

As is shown in Fig. 12, parameter $\mu_3^{(r)}$ does not seem very useful to predict the new tumor areas developed at day 5 (AUC = 0.67). These results were expected since many areas labeled as tumors at day 5, which have values of HbSat under the threshold for tumors at day 3, and are misclassified as normal. Table 4 shows the examples of selected areas classified as tumors at day 5 by the fluorescence maker but with low $\mu_3^{(r)}$ values at day 3. By examining the $\mu_t^{(r)}$ values in Table 4, we note a progressive increase in many regions, such as region 1911, where the $\mu_t^{(r)}$ value increases from 11.08 at day 1 to 19.9 at day 3. Considering only the reading at day 3 $\mu_3^{(r)}$ with a threshold value of 20, region 1911 would be misclassified as normal, but taking into consideration the increasing tendency of $\mu_t^{(r)}$ over time, region 1911 could be classified as tumor at day 3 (even though at day 3 it is still below the threshold for a tumor).

6.1 Delta Feature (δ)

In order to better classify regions as tumor/normal, it is necessary to evaluate the variations of the t values across different days. Since HbSat in tumor regions usually increases progressively, using cumulative values could help to better classify the regions. For any given time period $[t_s, t_f]$, and any region r

that has all its $\mu_t^{(r)}$ defined for that time period, we define the $\delta_{t_s \rightarrow t_f}^{(r)}$ (or simply δ) feature of the region as follows:

$$\delta_{t_s \rightarrow t_f}^{(r)} = \frac{\sum_{t=t_s}^{t_f-1} [\mu_{t+1}^{(r)} - \mu_t^{(r)}]}{\mu_{t_f-1}^{(r)}} \quad (3)$$

t_s = Starting day used for prediction; t_f = Final day used for prediction; r = Area of interest.

If any of the $\mu_t^{(r)}$ in $[t_s, t_f]$ are undefined, then $\delta_{t_s \rightarrow t_f}^{(r)}$ is undefined as well. In other words, any area missing a $\mu_t^{(r)}$ reading between T_s and T_f was discarded from this analysis. The δ feature for each region is stored in a matrix using day 1 as a reference for the label on each region as shown in Table 5.

Table 5 Example of δ features for regions 1688 to 1699. In this example $T_s = 1$ and $T_f = 3$. Any area missing a $\mu_t^{(r)}$ reading between days one and three is marked as NaN in all calculation of δ (e.g., areas 1688 and 1693). Column 3 [$\delta(2, 3)$] is used for forecasting the occurrence of tumor areas in day 5 in this example.

Region	$\delta(1, 2)$	$\delta(2, 3)$	$\delta(3, 4)$	$\delta(4, 5)$
1688	NaN	NaN	NaN	NaN
1689	1.141138	0.720068	0.806182	NaN
1690	1.325901	1.155489	NaN	NaN
1691	1.673515	0.878224	NaN	NaN
1692	0.45369	0.718235	0.70129	NaN
1693	NaN	NaN	NaN	NaN
1694	1.035799	0.835592	0.465785	NaN
1695	0.396392	0.717522	0.561913	0.429166
1696	0.787204	0.893385	0.419892	0.515459
1697	-0.12611	0.173589	0.616141	0.387803
1698	-0.30334	0.327984	0.484636	0.537067
1699	-0.07384	-0.40911	NaN	NaN

6.2 Forecasting Tumor Areas in Day Five using (δ) Feature

We assess the capability of δ features by forecasting the occurrence of tumor by conducting the following experiment. We evaluate 97 normal areas along with 176 tumor areas, which are classified as normal/tumor at day 5, with valid $\mu_t^{(r)}$ readings in days 1–3. The delta feature was used to classify the regions as tumor/normal at day 5. The $\delta_{1 \rightarrow 3}^{(r)}$ feature is used to classify regions as tumor/normal in day 5 by using a logistic regression model

$$\text{logit}(\text{odds-ratio of tumor in region } r \text{ at day 5}) \sim 1 + \delta_{1 \rightarrow 3}^{(r)}.$$

The resulting ROC curve is shown in Fig. 13(a).

Using the same methodology, we repeat the experiment in a different mouse. We evaluate 61 normal areas along with 52 tumor areas, which are classified as normal/tumor at day 5, with valid $\mu_t^{(r)}$ readings in days 1–3. The resulting ROC curve is shown in Fig. 13(b).

6.3 Cross Validation Study

The contribution of this paper is the potential to forecast newly developed tumor areas using values of HbSat over a course of several days before the formation of the tumor (the $\delta_{t_s \rightarrow t_f}^{(r)}$ feature). Since using the $\delta_{t_s \rightarrow t_f}^{(r)}$ feature is a novel approach, it is necessary to evaluate its performance through a cross-validation study. We conducted a 10-fold cross-validation experiment to assess the performance of the logistic regression-based classifier utilizing the $\delta_{1 \rightarrow 3}^{(r)}$ feature of each region r to forecast the label of region r at day 5. Combining the tumor and normal areas of the two different mice in Sec. 6.2, we use 228 tumor areas along with 158 normal areas. We partitioned the regions in a stratified manner at random into 10 disjoint sets of regions. We used each group of nine sets of regions as a training set to build the logistic regression model using the $\delta_{1 \rightarrow 3}^{(r)}$ features of the training set. Then, using the regression model, we predicted the labels of the regions in the validation set (the 10th set of regions). We found that the AUC of the ROC is 0.87. We concluded that our proposed δ feature is effective at forecasting future tumor/normal regions.

Fig. 13 Forecasting of tumor occurrence on day 5 using $\delta_{1 \rightarrow 3}^{(r)}$ feature at day 3 for all available regions in two different mice. ROC curve first mouse (a), ROC curve second mouse (b) (x-axis: false-positive rate; y-axis: true-positive rate).

7 Conclusion

Forecasting of newly developed tumor areas at day 5 using HbSat data from days 1 to 3 using the $\delta_{1 \rightarrow 3}^{(r)}$ feature gives AUC = 0.90 and AUC 0.85 in two different mice, and it is more effective than using the $\mu_3^{(r)}$ feature alone.

These results show that our proposed $\mu_t^{(r)}$ and δ features based on HbSat are useful to evaluate tumor progression and forecast new occurrences of tumor areas. We demonstrated the advantages of utilizing HbSat readings on multiple days to create an HbSat profile over using a single standalone HbSat reading for detection of newly developed tumor areas.

8 Discussion

This project was envisioned as a proof of principle to show that analyzing HbSat profile as a tumor microenvironment feature makes it possible to evaluate development of tumor progression over time in a mouse model. We used the “window chamber” model since many of the variables related to tumor growth are controlled, e.g., type of tumor cell, number of tumor cells, original location of the tumor, etc., and the ease with which microvessels can be imaged in this model. In the clinical setting, some or none of these variables are known and, therefore, in order to evaluate tumor progression further studies must be performed in more realistic models, such as standard medical *in vivo* image technology, e.g., CT or PET. For example, PET imaging can indicate tissue metabolic activity through regional glucose uptake, and using the δ feature developed herein, the glucose profile over time may give an indication of future tumor progression. Therefore, the methodology used in this paper can be used as a base for further studies with other imaging modalities. In future research, we plan to analyze PET images using the same methodology to evaluate cancer regression after therapy.

Acknowledgments

The data used in this study were collected in the laboratory of Mark W. Dewhirst (Duke University, Durham, North Carolina). The authors are grateful for ability to repurpose the data for use in this project.

References

1. X. Cheng et al., “Breast cancer detection by mapping hemoglobin concentration and oxygen saturation,” *Appl. Opt.* **42**(31), 6412–6421 (2003).
2. C. M. Carpenter et al., “Image-guided optical spectroscopy provides molecular-specific information in vivo: MRI-guided spectroscopy of breast cancer hemoglobin, water, and scatterer size,” *Opt. Lett.* **32**(8), 933–935 (2007).
3. A. J. Moy et al., “Wide-field functional imaging of blood flow and hemoglobin oxygen saturation in the roden dorsal window chamber,” *Microvasc. Res.* **82**(3), 199–209 (2011).
4. G. M. Palmer et al., “Monte Carlo-based inverse model for calculating tissue optical properties. Part II: application to breast cancer diagnosis,” *Appl. Opt.* **45**(5), 1072–1078 (2006).
5. B. Brooksby et al., “Imaging breast adipose and fibroglandular tissue molecular signatures by hybrid MRI-guided near-infrared spectral tomography,” *PNAS* **103**(23), 8828–8833 (2006).
6. B. Chance et al., “Breast cancer detection based on incremental biochemical and physiological properties of breast cancer: a six-year, two-site study,” *Acad. Radiol.* **12**(8), 925–933 (2005).
7. B. S. Sorg et al., “Hyperspectral imaging of hemoglobin saturation in tumor microvasculature and tumor hypoxia development,” *J. Biomed. Opt.* **10**(4), 044004 (2005).
8. B. J. Moeller et al., “Radiation activates HIF-1 to regulate vascular radiosensitivity in tumors: role of reoxygenation, free radicals, and stress granules,” *Cancer Cell* **5**(5), 429–441 (2004).

Miguel R. Ossandon received his BS degree in medical technology from the University of Tarapaca, Chile, and his MS degree in computer science from George Washington University, Washington, DC. He is currently a PhD student at the University of Maryland, Baltimore County, and a program director at the National Cancer Institute in the Technology Development, Brach. His research interests include computational modeling for cancer research and medical imaging processing.

Dhananjay S. Phatak earned his PhD in computer engineering from the University of Massachusetts (UMASS), Amherst, in 1994. He received a BTech degree in electrical engineering from the Indian Institute of Technology (IIT), Mumbai (Bombay), India, in 1985. Currently, he is a faculty of computer science at the University of Maryland, Baltimore County (UMBC) in the CSEE Department. His main research interests currently include all aspects of cyber-security; computer arithmetic algorithms; number theory; cryptology and bioinformatics.

Brian S. Sorg (SM'11) received his BS in electrical engineering from the University of Maryland, and an MS and a PhD in biomedical engineering from Johns Hopkins University and the University of Texas at Austin, respectively. Previously, he was a research associate at Duke University Medical Center, and assistant professor of biomedical engineering at the University of Florida. He is currently a program director at the National Cancer Institute. Research interests include applications of biophotonics to cancer.

Konstantinos Kalpakis received his PhD degree in computer science in 1994. He is currently an associate professor at the Department of Computer Science and Electrical Engineering at the University of Maryland, Baltimore County. His research interests include wireless sensor networks, query processing in distributed systems, and mining time series datasets. He is a senior member of the IEEE Computer Society.