

Computer Aided Diagnostics in Digital Pathology: Automated Evaluation of Early Phase Pancreatic Cancer in Mice

Leeor Langer · Yoav Binenbaum · Leonid Gugel · Moran Amit · Ziv
Gil · Shai Dekel

Received: date / Accepted: date

Abstract

Purpose: Digital pathology diagnostics are often based on subjective qualitative measures. A murine model of early phase pancreatic ductal adenocarcinoma provides a controlled environment with a priori knowledge of the genetic mutation and stage of the disease. Use of this model enables the application of supervised learning methods to digital pathology. A computerized diagnostics system for histological detection of pancreatic adenocarcinoma was developed and tested.

Methods: Pathological H&E stained specimens with early pancreatic lesions were identified and evaluated with a system that models cancer detection using a top-down object learning paradigm, mimicking the way a pathologist learns. First, the dominant primitives were iden-

tified and segmented in the images, i.e. the ducts, nuclei and tumor stroma. A boost based machine learning technique was used for duct segmentation, classification and outlier pruning. Second, a set of morphological features traditionally used for cancer diagnosis which provides quantitative image features was employed to quantify subtle findings such as duct deformation and nuclei malformations. Finally, a visually interpretable predictive model was trained to distinguish between normal tissue and pre-malignant cancer lesions, given ground truth samples.

Results: A predictive success rate of 92% was achieved using 10-fold cross validation and 93% on an independent test set. Comparison was made with state of the art classification algorithms which are not interpretable as visible features yielded the contribution of individual primitive features to the prediction outcome.

Conclusions: Quantitative image analysis and classification was successful in pre-clinical histology diagnosis for early stage pancreatic adenocarcinoma. These methods may be applicable to a diverse range of malignancies. The use of interpretable supervised learning techniques may improve the success of CAD in histopathological diagnosis.

Leeor Langer
School of Mathematical Sciences, Tel Aviv University, 69978
Tel Aviv, Israel
Tel.: +972-49-836110
E-mail: leor.langer@gmail.com

Yoav Binenbaum · Moran Amit
Laboratory for Applied Cancer Research, Rambam Medical
Center E-mail: y_binenbaum@rambam.health.gov.il,
m_amit@rambam.health.gov.il

Leonid Gugel
Computer Science department, Technion Israel In-
stitute of Technology, 32000 Haifa, Israel E-mail:
leong@cs.technion.ac.il

Ziv Gil
Rappaport School of Medicine, Technion Israel Institute
of Technology and the Laboratory for Applied Cancer
Research, Rambam Medical Center, 31096 Haifa, Israel
E-mail: g_ziv@rambam.health.gov.il

Shai Dekel
General Electric Global Research and the School of Mathe-
matical Sciences, Tel Aviv University, 69978 Tel Aviv, Israel
E-mail: shai.dekel@ge.com

Keywords Cancer prediction · Boosting · Segmenta-
tion · CAD

1 Introduction

The aim of this work is the development of a comput-
erized diagnostics tool which can assist pathologists in
decision making when confronted with the elusive diag-
nosis of early stage Pancreatic Ductal Adenocarcinoma
(PanIN). We propose and evaluate a cancer prediction

system for early phase pancreatic ductal adenocarcinoma in a murine mice model. Such a *controlled environment*, in which the genetic mutation and stage of disease evolution are known a-priori, is a classic setting for the usage of supervised learning methods. The advantage of such methods is that concepts (such as ductal morphology) can be defined and represented without the need for explicit programming, but rather by being presented with previous experiences [23]. Adapting such a system to a human clinical setting can be done using representative training samples, assuming the concepts do not change. When applied to human cases, the system would require adaptation yet no conceptual change.

Pancreatic Ductal Adenocarcinoma (PDAC) is the fourth leading cause of cancer death in the US, with an estimated incidence of 37,700 cases and 34,300 deaths from the disease in 2008 [15]. Like other epithelial neoplasms, pancreatic cancer starts as a curable noninvasive lesion termed Pancreatic Intraepithelial Neoplasm (PanIN). Detection of this early precursor of cancer opens a window of opportunity for cure. Despite its aggressive nature, diagnosis of pancreatic cancer poses a great challenge to pathologists. This is because low grade cancer shares many of its morphological features with those seen in chronic inflammatory conditions of the pancreas. Correct differentiation of PanINs from benign conditions on one hand and from invasive cancer on the other, offers the best hope for treating pancreatic cancer before an incurable disease has developed. Using the toolbox available for pathologists today, diagnosis of PanIN is extremely difficult and sometimes even impossible [17].

The *novelty* of our system is twofold. First, we introduce a duct segmentation technique based on unique features and supervised learning methods with outlier pruning. The objective of this method is to quantify the concept of subtle morphological changes in ducts, due to their clinical significance in PDAC assessment. Second, our CAD system is open and interpretable. We generalize the “second opinion” CAD systems [6] which only provide a classification result by incorporating feature analysis into an interpretable classifier output. Contrary to other CADs, our system does not emphasize numerous features coupled with dimensionality reduction techniques, instead we base our classification model on a small number of clinically inspired features, akin to the visual inspection performed by a physician.

In order to pathologically diagnose lesions in the pancreas, a specimen of the pancreatic tissue is preserved with chemical fixatives and processed to 5 micron thick layers that are mounted on glass slides. The

slides are then stained with Hematoxylin and Eosin (H&E), the gold standard stain in histology which colors the cellular organelles blue and pink. Routinely, slides are then manually examined under a light microscope by the pathologist [10]. Several features can be used to distinguish PanINs from different conditions affecting the pancreas, and most of these boil down to the violation of tissue architecture by cancerous lesions as opposed to benign conditions. In neoplastic states the glands violate the lobular architecture; the glands can invade arteries and nerves; The stroma, the connective tissue which forms the supportive framework for the glands, is abundant and dense; The lumen of the ducts or the glands themselves may be incomplete; and the size of the cell’s nuclei vary by a ratio greater than 4:1 [17]. A major obstacle to the correct diagnosis of PanIN stems from the fact that all but the last of these features are *descriptive, subjective to assessment and hard to quantify*. Therefore, automatic quantification of some of these features may be beneficial to pathologists.

In recent years advances in virtual microscopy in general and whole slide imaging in particular have opened the door for computerized diagnostics [25]. Despite such advances, pathology has exhibited a low adoption rate for digital workflows compared to radiology. One possible explanation is that the sheer size of the data from this modality poses challenges to all computerized diagnostics systems. A high resolution chest scan is comprised of approximately 134 million voxels, while a single prostate biopsy procedure with 12-20 samples digitized at $\times 40$ resolution encompasses a dataset of 2.5-4 billion voxels [12].

In order to deal with such challenges, research in digital pathology has focused not only on computer aided diagnostics but also on specific object level segmentation techniques. Various approaches have been proposed for nuclei segmentation, such as fuzzy logic based segmentation [2], template matching schemes [18] and adaptive thresholding [20]. Segmentation algorithms for other structures such as glands have also been proposed, such as a segmentation algorithm based on active contours [1] and pixel level boundary classification method [18].

An overview of the challenges and CAD systems in histology is presented in [14]. One such system suggested by Doyle et. al. provides an automated Gleason grading system for prostate cancer based on graph, morphological, and textural features together with an SVM classification framework [8]. Additional research on prostate cancer classification using a multi-resolution framework which identifies cancerous regions at a low resolution and consequently examines and classifies such regions in greater detail was introduced in [7]. Gurcan

et. al. have introduced analysis methods of neuroblastoma [13], focusing on nuclei segmentation based on top hat reconstruction and hysteresis thresholding. Additional work on neuroblastoma computer aided diagnostics was introduced in [26], utilizing a classification framework based on texture features extracted using co-occurrence statistics and local binary patterns. Roullier et al. have focused on breast cancer using unsupervised learning methods together with a multi-resolution framework and domain specific knowledge for mitosis extraction and visualization [24].

Contrary to the above examples, our system defines clinical aspects and feature extraction is carried out based on such definitions and not on mathematically inspired models. Our CAD implements such a strategy by analyzing unique features of PDAC, based on morphological object analysis. First, key clinical primitives are identified, i.e. ducts, nuclei and stroma. A boosted strong classifier is trained based on manually annotated examples in order to distinguish between duct contours and non-duct contours. Nuclei and stroma are segmented using unsupervised methods. Second, object level features based on clinical aspects are extracted from the identified primitives. Finally, a classification tree is trained and evaluated based on the ground truth samples. Since such a classifier is easily interpretable and all features are quantifiable the output of such a predictive model can be used as a decision support “flowchart”.

As PDAC stems from the ductal cells, we postulated that the ability to characterize and analytically quantify duct morphology may play a pivotal role in the prediction of early phase cancer. Novel features and training methods are introduced, aiming to address the properties of this unique region of interest. The features of pixels on the duct contour are viewed as a sequence of values collected over the boundary curve whereby successive values are related to previous ones, forming a stochastic process in the feature space. This approach is utilized to remove outliers during model creation and to extract object level features.

This paper is organized as follows. A detailed description of the dataset used in this study is described in section 2. An outline of the algorithm is depicted in section 3. A description of H&E color based preprocessing is presented in section 4. We introduce duct detection including feature extraction, training, testing and evaluation in section 5. Nuclei and stroma detection are outlined in section 6. In section 7 object level features of the primitives are detailed. We discuss image level classification and results in section 8. Conclusions and future work are explored in the final section.

2 Materials

2.1 Pancreatic Tissues

In this study image acquisition of pancreatic specimens was carried out using a whole slide scanner. Ground truth samples from cancer induced mice are manually examined in a traditional pathologic fashion, grouping areas of normal pancreatic histology together with inflammatory areas, as areas displaying PanIN are grouped separately. Selected regions of interest from both groups are then studied by the system for analysis and classification.

In order to take advantage of a controlled environment, we used a transgenic mouse model of PDAC that was developed by Tuveson et al [16]. Two culprit genetic mutations that exist in the majority of PDAC cases in humans were expressed in mice pancreata. Mice carrying these two mutations concomitantly develop PDAC that recapitulates the genetic makeup and pathological characteristics of human PDACs [5], [19]. Moreover, the natural history of the disease in mice is highly predictable: mice are born with normal pancreas; by the age of 10 weeks the majority of mice display PanIN lesions; by 5 month an invasive tumor is evident and a metastatic spread soon follows.

For this study we used pancreatic specimens from four mice at the ages of 4 weeks, and four mice at the age of 10 weeks. Samples were fixed, processed in a single batch and stained with H&E. Each pancreatic specimen yielded 4 slides representing different tissue depths. Tissue processing was done based on the routine pathology lab workflow, according to traditional guidelines. All experimental procedures were performed under the auspices of our approved animal protocol in accordance to the guidelines issued by the Institutional Animal Care and Use Committee.

2.2 Microscopy

Digital image acquisition with the aid of whole slide scanners has been rapidly adopted by pathologists in the last decade, due to its ability to produce a high resolution photomicrograph of an entire histologic slide. Whole slide images (WSI) have several advantages over conventional microscopy: WSI can be explored at different magnifications and in-between magnifications; quality of WSI is constant over time, focusing is carried out during scanning and images of multiple focal planes enable continuous accurate focus over the entire slide. In this work we used WSI technology (Panoramic MIDI, 3DHitech, Hungary) in order to maximize the reproducibility of the images captured in terms of image

quality, focus, lighting, color variations and optical configuration. The scanner is equipped with 3CCD progressive scan color camera with a resolution of 1360H and 1024V pixels (HV-F22, Hitachi Kokusai Electric, Japan). Slides were scanned at $\times 45$ magnification and viewed on a Panoramic Viewer platform (3DHitech, Hungary). Output file format was lossless tif.

A database of whole slide images of pancreatic specimens stained with H&E was acquired. Pathological classification of specimens was carried out by two physicians independently, which were blinded with respect to tissue origin. Regions of interest (ROIs) were annotated by each physician independently and include ducts, focuses of inflammation, acini and areas of stroma. Discrepancies between the two were resolved by discussion and consensus.

A database of 103 images of 1211x1044 pixels was acquired and grouped into two series; 43 ROIs displayed healthy and inflammatory zones, and a second group of 60 ROIs of displayed PanINs. 80% of the ROIs in the database include images of ducts, while 20% did not, focusing on inflammatory zones, acini and stroma. Moreover, an independent test set comprised of 30 images is used for evaluating our system. 6 additional images were taken to be used as a training set for duct detection.

3 Algorithm Overview

A complete outline of our CAD algorithm is as follows: Given a new input image, we first preprocess the image by attributing each pixel as belonging to one of three color based clusters; lumen, cellular component and stroma. Second, we identify and segment the key structures in the image; ducts, nuclei and stroma. Ducts are identified and segmented using supervised learning, based on manually annotated contour examples, the nuclei and stroma are identified and segmented using unsupervised methods. Third, we extract features from the identified primitives which are inspired by clinical definitions. Finally, we classify the image as either healthy or PanIn based on a decision tree classifier.

The decision tree, together with the object based features, form a flowchart which can be used as a report for early cancer classification. A physician can inspect the tree from the root and examine each split until reaching a leaf which contains the final decision. This enables the physician to make a *controlled decision* based on *quantitative measurements* and to individually evaluate different components of the decision making process. Note that our system contains four tiers: preprocessing, detection, classification and explanation.

Note that the more challenging tasks are performed using supervised learning methods.

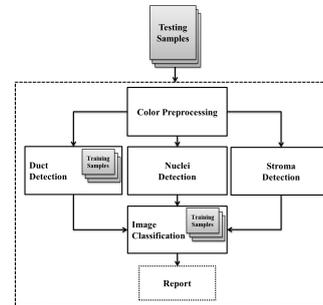


Fig. 1 System Block Diagram

4 Color Preprocessing

H&E staining is a technique used to highlight specific entities in a tissue specimen using colored dyes. Hematoxylin is a purple dye which stains the chromatin within the nuclei, while Eosin is a pink dye which stains the cytoplasm and connective tissue. As a preprocessing step to H&E object recognition, each pixel belonging to an individual testing image is transformed from its original RGB value into a four dimensional color descriptor. Such a descriptor decouples intensity information from color information in the image. The first two component of the pixel color descriptor are the a^* and b^* color components from the CIE-Lab color-space image transform. Uniform changes in such components correspond to uniform changes in human perceived color [9], contrary to the RGB color space which models the output of physical devices and not the nonlinear response of the human eye.

The second two components of the pixel color descriptor are the contribution of each individual stain to a pixel. The problem of decomposing an image into separate dyes in an unsupervised manner [22] can be formulated as the following matrix factorizations problem:

$$\operatorname{argmin}_{D,S} \|I - D \cdot S\|_{L_2} \quad (1)$$

subject to $D, S \succ 0$

In the above formulation I is a $3 \times N$ matrix, where N is the number of pixels in a testing image, D is a 3×2 matrix where each column corresponds to one of the two dyes used in the staining process and S is a $2 \times N$ matrix such that each entry $S_{i,j}$ indicates the contribution of dye i to pixel j .

The color descriptor is then defined by a 4-dimensional vector representing the color signature of a pixel, consisting of the CIE-Lab color components and the contribution of each dye to the pixel

$$p_j = [a_j^*, b_j^*, S_{1,j}, S_{2,j}] \quad (2)$$

The motivation behind such a color descriptor, as opposed to simply clustering a three dimensional signature consisting of the RGB values, is to introduce as much prior knowledge as possible to our system. The CIE-Lab colorspace was introduced to simulate human vision such that distances in CIE-Lab correlate with perception of color. As the manner by which Hematoxylin and Eosin stain different cellular organelles and tissue components is known, deconvolution into these two colorspace enables integration of such prior knowledge to our system (i.e. nuclei stain purple while cytoplasm stains pink). Such approaches were successfully introduced into various systems in the field of digital pathology, such as [27].

We then proceed by applying 3-means clustering on the color descriptors so as to partition the image into discriminative color groups, which will give us a “first order approximation” of the image primitives. A choice of $k = 3$ stems from the fact that each image contains 3 distinct structures, i.e. cellular component (with nuclei), stroma and lumen. During early stages of cancer clustering is a viable technique for object detection since all primitives are still well-formed and each structure consists of a unique color composition. This is in contrast with object recognition in more advanced cancer, whereby structures lose their architecture and color composition no longer describes the primitives. This approach overcomes some of the challenges related to color clustering in H&E staining [25] where using color representations based on only two dimensions results in poor object recognition. By fusing color information from both color space transform and color deconvolution techniques, distinct clusters in the color description space are formed

5 Duct Detection

5.1 Duct Classifier Construction

The goal of the duct classification is to find a good local approximation of the duct curves (if such curves exist) and the local sets of interior and exterior points. Following color preprocessing, a set of candidate contours are calculated. Since ducts are structures that form channels, circumscribed by nuclei, the external boundaries of each lumen cluster connected component region are used as a set of candidate contours.

To this end, we use a classifier that was trained on a set of features encoding the domain specific knowledge of such structures. It is trained based on manually annotated contours, using the classical formulation of Adaboost [11]. A strong binary classifier H is constructed from a set of weak classifiers, forming a linear combination of weighted hypotheses

First, each member pixel x_i in a candidate contour is classified by the strong Adaboost classifier

$$H(x_i) = \sum_{t=1}^T \alpha_t h_t(x_i) \quad (3)$$

where $\{h_t\}$ are the weak classifiers. An entire contour is classified as a duct if the total score of the entire contour is positive

$$H_{Duct}(\{x_i\}) = \text{sign}\left(\sum_{i=1}^N H(x_i)\right) = \text{sign}\left(\sum_{i=1}^N \sum_{t=1}^T \alpha_t h_t(x_i)\right) > 0 \quad (4)$$

We now review the underlying features used for classification. For each candidate pixel x_i , we calculate an optimal bisection of the box Ω^i , of size 60×60 pixels centered at x_i into *exterior* and *interior* regions. This is achieved by minimizing

$$\sum_{x_j \in \Omega_{interior}} (I(x_j) - 1) + \sum_{x_j \in \Omega_{exterior}} (I(x_j) - 0) \quad (5)$$

over all bisections $\Omega_{interior} \cup \Omega_{exterior}$ by lines going through x_i with angles discretized to multiples of 45° with I representing a binary mask with lumen pixels belonging to a specific duct are attributed a value of 1 and all other pixels are set to 0.

Once the optimal bisection is computed, we can proceed to calculate a set of features associated with the pixel:

1. Percentage of exterior region pixels belonging to nuclei cluster.
2. Percentage of interior region pixels belonging to lumen cluster.
3. Fixative Ratio: percentage of entire contour interior pixels (not only box interior) which do not belong to the lumen cluster.
- 4-9. Sum of squared DCT coefficients belonging to the six subbands of the two highest frequencies.
10. Deviation from average radius. The radius of a pixel is defined as the distance from the shape centroid. The average radius is the mean radius of all individual pixels belonging to a candidate contour. This

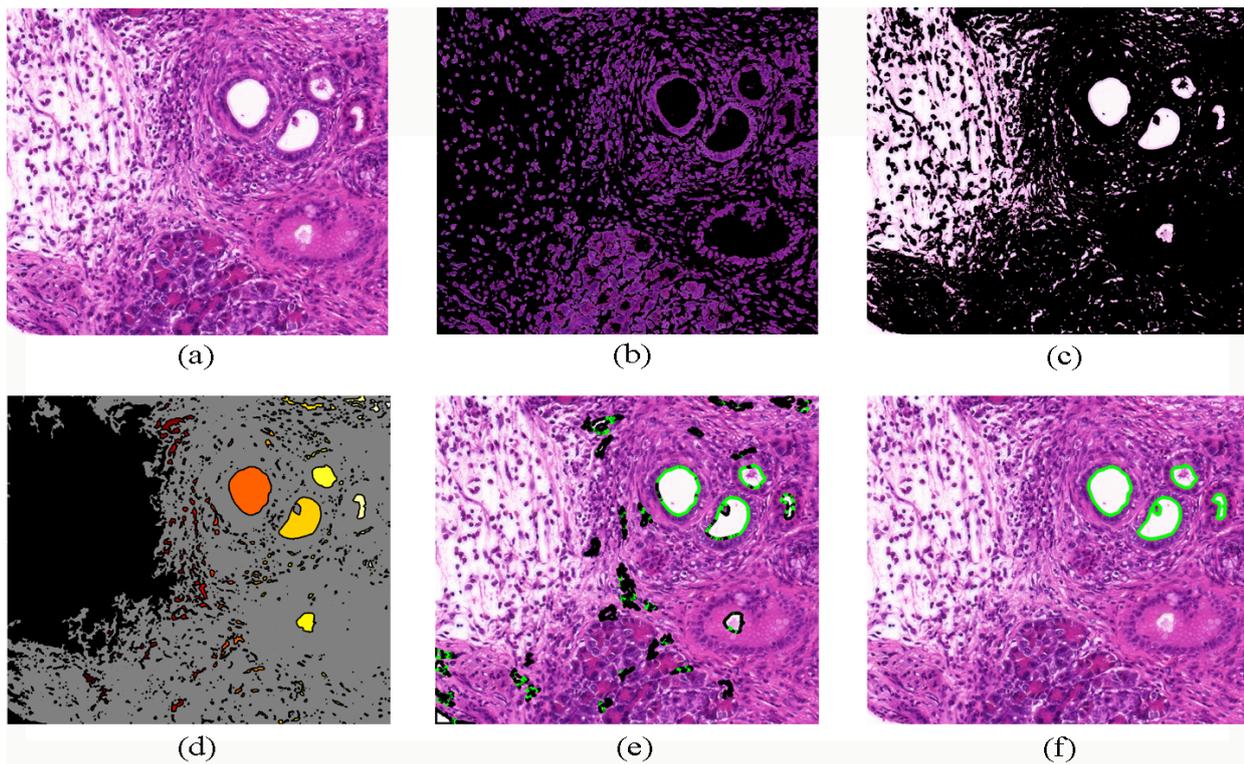


Fig. 2 Illustration of duct detection (a) Original image, (b) Nuclei cellular component color cluster, (c) Lumen cluster, (d) Candidate regions for detection, (e) Boundary classification results, green pixels are predicted as belonging to duct boundary, black pixels are predicted as not belonging to duct boundary, (f) Final result of duct identification

feature is calculated as $R(i) = \frac{|r(i) - \hat{r}|}{\hat{r}}$ where $r(i)$ represents the i^{th} pixel of the candidate contour and \hat{r} represents the average radius.

11. Deviation from perpendicular radius. The perpendicular radius is defined as the radius of the pixel which lies at a 90° angle to the i^{th} pixel of the candidate contour. This feature is calculated as follows $R(i) = \frac{|r(i) - r_p(i)|}{\hat{r}}$ where $r_p(i)$ is the pixel which lies at a 90° angle to the i^{th} pixel of the candidate contour.
12. Average contour curvature. A polygonal approximation of all discrete points on the contour is computed, followed by the calculation of curvature. The mean is taken over all values of the curvature.

The first three features measure the extent to which a pixel belongs to a setting befitting a duct, i.e. lumen interior, nuclei exterior and fixative within the duct. The fourth to ninth features determine the unique texture of the nuclei. DCT coefficients are an efficient texture classification feature and can represent a texture “signature”. The tenth feature models contour deformities, favoring smooth contours over contours with malformations. The eleventh feature favors circular contours over elliptic, in order to separate between ducts and artifacts caused during tissue processing. The last

feature favors contours with lower curvature distinguishing between ellipses and random angular structures.

Training the duct classifier requires a collections of manually annotated contours, with examples of ducts and examples of other closed contours which are not ducts (tears in the tissue, connective tissue closed contours and other structures). However, such a training set contains a substantial amount of outliers. The reason for this is that all ducts inherently contain imperfections, such as gaps in the surrounding nuclei, inaccuracies in the textural patterns of the surrounding nuclei due to limitations of WSI focus and small deformations in the shape of the duct. Since such imperfections tend to concentrate in specific areas of a duct (see Fig 3(d)), we can design an algorithm to identify and remove outlier boundary pixels.

We approach the problem of training a classifier on a noisy dataset by taking advantage of the temporality of neighboring pixels on a boundary contour, as visualized in Fig 3(c). We define the neighborhood of a pixel as:

$$\rho(x_i) = [x_{i-\delta}, x_{i+\delta}] \quad (6)$$

After an initial estimated classification of the noisy training set, we define $c_i = 1$ if the classification of pixel x_i is correct and $c_i = -1$ if the classification of x_i

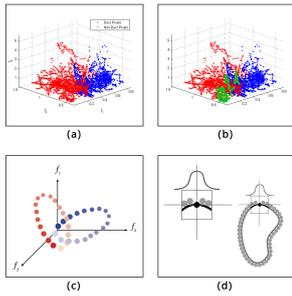


Fig. 3 Training the duct classifier (a) Feature space of entire hand marked training collection using three most dominant features of the boosted ensemble, (b) Removal of non-representative outliers, marked green, (c) Visualization of spatially neighboring pixels in the feature space, (d) Visualization of duct imperfection concentrated in a specific spatial region.

is incorrect. The region classification error of a pixel is defined as:

$$\epsilon_S = \sum_{x_i \in \rho(x_i)} c_j \cdot G(x_j) = \sum_{j-\delta}^{j+\delta} c_j \cdot G(x_j) \quad (7)$$

$$G(x_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (8)$$

In our experiment we used $\delta = 15$, $\sigma = 15$. The region classification error aggregates the classification error of a pixel together with its neighbors. The Gaussian weights emphasize classification of *spatially* close pixels within the specified neighborhood. A pixel with high ϵ_S represents a neighborhood $\rho(x_i)$ with high classification confidence. Pixels with low ϵ_S represent a neighborhood $\rho(x_i)$ with low classification confidence.

Regions with low classification confidence are in effect confusing training samples. Removing such samples aids in alleviating boosting bias [28] while taking advantage of the temporal properties of the data. Since AdaBoost tends to focus on the hard-to-classify samples, removing such samples which belong to a region which is not representative of a duct will reduce the effects of over-fitting. We define the outliers using the context of the classification problem, taking into account a pixel along with its neighborhood in order to determine non-representative samples.

5.2 Duct classifier evaluation

We evaluate the performance of our method using 10-fold cross validation. The duct classifier was trained on a

sample size of 10,000 observations with 12 features. We examine the predictive capabilities of a classifier trained on the entire noisy dataset and the classifier trained on the pruned dataset, after removing the confusing samples. Note that in order to compare both methods, training is performed on the pruned training set yet testing the pruned classifier is done on the same folds as that of the complete training set, without removing the outliers.

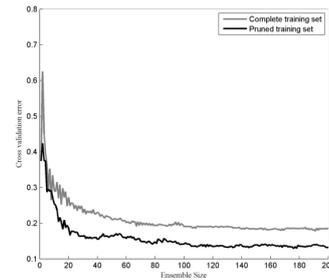


Fig. 4 Duct detection evaluation using 10-fold cross validation

The classifier trained on the pruned dataset exhibits superior performance, with cross validation error converging to 15%. The classifier trained on the complete dataset does not perform as well, converging to an error of 21.5%. These results, as shown in Fig 4, illustrate the importance of pruning outliers.

6 Nuclei and Stroma Detection

6.1 Nuclei Segmentation

The initial color based clustering is not accurate enough for nuclei segmentation, since nuclei exist inside and outside pancreatic cellular components. In order to refine the nuclei segmentation, a marker controlled watershed transform [3], [21] method is used in order to take advantage of image intensity and gradient, not only color. Since this method is sensitive to the choice of initial markers for the flooding algorithm, additional morphological operations are utilized.

Initial markers or “basins” for the flooding algorithm are selected by taking connected component regions belonging to the nuclei\cellular component cluster. The watershed is initiated from such pixels or markers whose connected component area is larger than $area_{min}$ and is smaller than $area_{max}$. These constraints are set so as to avoid large tissue areas and small groups of pixels. Large areas belong to tissue regions and small areas are in effect noise, such pixels usually do not depict good candidates for nuclei segmentation.

A candidate basin is therefore a set of pixels which belong to the nuclei\cellular component cluster and conforms to the connected component limitation, i.e.

$area_{min} < area < area_{max}$. The flooding algorithm begins with such markers and terminates at a local minimum of the gradient of the CIE-Lab image intensity ∇L or by reaching the local minima imposed by the geodesic zone of influence of a candidate basin\ nuclei. Thus, image intensity is fused with color analysis to achieve a refined nuclei segmentation.

6.2 Stroma Segmentation

Since the stroma has a unique color composition which is faithfully digitized by a high quality whole image scanner, a “rough” segmentation using the color clustering techniques detailed above yields favorable results. The stroma segmentation is therefore simply denoted as all pixels belonging to the stroma cluster, as detailed in section 4. Since the features extracted from the stroma are fairly simple, this segmentation, coupled with feature extraction from other histologic objects, is sufficient to properly characterize an image.

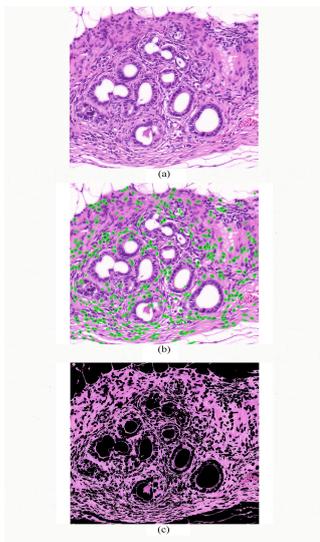


Fig. 5 Nuclei and stroma segmentation examples (a) original image, (b) segmented nuclei marked in green, (c) segmented stroma

7 Object Level Feature Extraction

After identification and detection of pancreatic primitives, a set of features are extracted from each primitive. Duct deformations and imperfections are modeled

by taking a detailed approach of individual pixels and a holistic view of the entire duct region. Since AdaBoost allows interpretation of importance of each weak classifier, the most important feature from the duct detection phase (ratio of nuclei pixels in box exterior 1) is derived. The values of such a feature over the contour can be viewed as a stochastic process $X(t)$, which best describes the evolution of the classifier over boundary (see Fig 3(c)). Using this formulation, each pixel feature represents a discrete time variable with index t_i corresponding to pixel x_i on the contour.

The following features are extracted from *duct* objects:

1. Mean Drift energy $E|S_{\frac{1}{3}}[X(t)]|$. Duct imperfections are modeled as *drift* in the spectrum. $X(t)$ represents the most important feature of the duct classifier. $S_{\frac{1}{3}}$ represents the spectrum $S[(X(t))]$ multiplied by a lowpass filter with bandwidth equaling a third of the entire spectrum.
2. Mean Imperfection ratio: The ratio of the duct classifier negative and positive scores $\frac{\sum_{i=0}^N H(x_i) < 0}{\sum_{i=0}^N H(x_i) > 0}$. The “false positives” of the classifier are indicative of duct deformity.
3. Mean duct convex area.
4. Mean duct eccentricity, the ratio of the distance between the foci of the enclosing ellipse and its major axis length.
5. Number of segmented ducts.

The following features are extracted from *stroma* object:

1. Stroma “mass”: Number of pixels belonging to stroma object.
2. Stroma “haphazard” measure: Solidity of connected component regions, the ratio of pixels belonging to the convex hull of each connected component region.

The following features are extracted from *nuclei* objects:

1. Mean area.
2. Extra-nuclei intensity variance, the standard deviation of CIE-Lab intensity L taken over the entire nuclei collection $\sigma(L)$.
3. Extra-nuclei color variance, standard deviation of CIE-Lab color $C = \sqrt{a^* + b^*}$ taken over the entire nuclei collection $\sigma(C)$.
4. Intra-nuclei intensity variance: Each nuclei intensity variance is calculated separately. Defined as $max(\{\sigma(L_i)\}_{i=1}^{N_{nuclei}}) - min(\{\sigma(L_i)\}_{i=1}^{N_{nuclei}})$ where N_{nuclei} denotes the number of segmented nuclei.

5. Intra-nuclei color variance: Each nuclei color variance is calculated separately. Defined as $\max(\{\sigma(C_i)\}_{i=1}^{N_{nuclei}}) - \min(\{\sigma(C_i)\}_{i=1}^{N_{nuclei}})$.

The aforementioned features are *not* the same as features defined in the clinical literature. For example, such literature defines 4:1 ratio as a quantifiable measure for cancer grading [17]. From a statistical standpoint, such a measure is prone to noise, taking into account only 2 nuclei areas (the maximal and minimal) instead of a more robust statistic. Other features are suggested that are more robust yet still capture the essence of nuclei anamorphosis and other clinical aspects.

8 Results and Discussion

8.1 Performance Evaluation and Analysis

Ultimately our method has two goals. First, it needs to appropriately represent and generalize the problem by achieving favorable results for a given testing criteria. Second, it needs to be easily interpretable in order to be incorporated into a physicians workflow. We compare the results of a single decision tree compared to a bagged ensemble (Random ForestTM) and a boosted ensemble (Boosted Decision Tree). On the one hand, a decision tree is highly interpretable and the pathologist finds this decision support method useful. On the other hand, ensembles of decision trees are considered state of the art with respect to performance [4].

We compare between a decision tree classifier and state of the art ensemble methods using 10 fold cross validation. We compared the prediction success rate, sensitivity and specificity of a single pruned decision tree, a bagged ensemble and a boosted ensemble. The decision tree based model functions worst, with a 92% across all three parameters. The Random Forest has the highest sensitivity (96%) compared to a boosted decision tree which exhibits the highest specificity (98%).

Another performance evaluation of our CAD system is an evaluation based on an independent test set. We trained a decision tree based on the entire aforementioned dataset. This model was used to predict the state of 30 images taken from 5 different slides. These slides were extracted from murine mice samples which have the same genetic mutation and are annotated in the same manner as the training set. Our CAD exhibited a 93.3% predictive success rate. The specificity on this set is 100%, while the sensitivity is 86.7%.

In order to evaluate the impact of the individual primitives on the result outcome, we compare decision trees grown based on their corresponding features using

10 fold cross validation, as depicted in Fig 6. The duct based model performs best with respect to prediction success rate and sensitivity, yet worst with respect to specificity. This model is based on features which describe duct deformation; therefore it can better depict early cancer compared to healthy observations. This corresponds to our hypothesis that because pancreatic cancer originates at the ducts, slight variations in ductal morphology might serve as an early predictor of cancer.

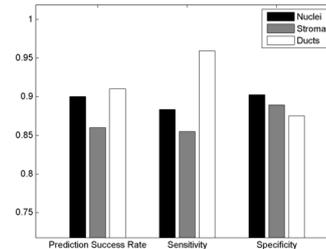


Fig. 6 Feature comparison using 10-fold cross validation, grouped by primitive

Duct analysis is an important aspect of our CAD system. Note that most duct features do not necessarily rely on a successful segmentation, see duct features list 1. We provide a methodological outline of duct analysis in Fig 7 with an emphasis on duct segmentation using contour pixel classification and aggregation. Contour pixels marked green are pixels classified as belonging to a duct boundary, black pixels are classified as not belonging to a duct boundary. Score aggregation over the entire contour is described in equation 4.

Subfigure (a) of Fig 7 is an example of a PanIN image. ROIs are marked according to the aggregated classification score (true positive, true negative, false positive and false negative). Note that malformations across the duct contours are marked black. This phenomenon becomes especially pronounced in the ROI marked FN. This duct is not successfully classified as a duct due to the amount of its malformations, although features are extracted from this contour nonetheless.

The ROI marked TP is an example of a successful duct segmentation, note that the malformation in the bottom side are marked black. The ROI marked FP is an example of a gap in the connective tissue surrounded by nuclei which is incorrectly classified as a duct (false positive). The ROI marked TN is an example of a candidate contour which is correctly classified as non duct structures (true negative).

We note that various applications can benefit from "tuning" duct classification, for example to correctly classify a false negative as marked in the ROI FN. This

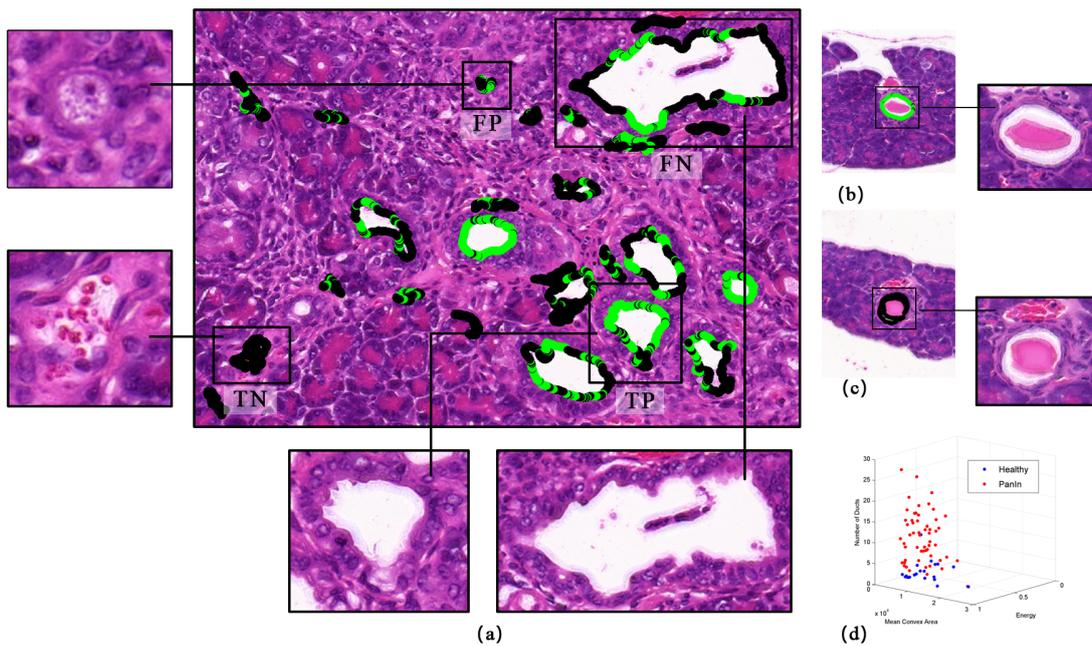


Fig. 7 Illustration of duct segmentation and classification (a) PanIn image with ROIs; includes TP,FP,TN and FN of duct contour classification,(b) Healthy image with a well formed duct,(c) Healthy image with fixative filled duct,(d) Duct feature distribution

can be achieved by generalizing equation 4:

$$H_{Duct}(\{x_i\}) = \text{sign}\left(\sum_{i=1}^N \sum_{t=1}^T \alpha_t h_t(x_i)\right) > \beta \quad (9)$$

In this research we used $\beta = 0$. A choice of $\beta < 0$ will reduce the number of false negatives at the expense of increasing the number of false positives. A choice of $\beta > 0$ will have the opposite effect.

Subfigure (b) of Fig 7 is an example of a healthy image. This image contains an ROI of a well formed duct. Note that almost all the pixels of its contour are successfully marked as green (i.e. belong to a duct contour). Subfigure (c) of Fig 7 is another example of a healthy image. The ROI in this image is an example of a failure to correctly classify a duct structure due to inaccurate candidate contour generation. Duct identification is difficult when the lumen is filled with fixative material up to its border, since candidate contours are generated by lumen connected component regions.

Subfigure (d) of Fig 7 is the distribution of 3 duct features across our training set. Note the formation of clusters in the feature space, which is of high importance to our prediction model. Most points falling at the origin are simply images whose ductal morphology did not play a part, such as inflammatory regions.

The limitations of our system are mainly the object level segmentation. Adjacent nuclei will not be correctly segmented by the watershed transform. Such a scenario poses difficulties for any nuclei segmentation algorithm, since no clear boundary exists. The stroma segmentation will sometimes include pixels that do not belong to

the stroma, since the segmentation relies on color clustering. Duct identification is most difficult in scenarios abundant with adjacent nuclei, forming duct-like structures or when material fills the lumen which will influence duct contour formation. We partially overcome the last problem by using morphological closing over the lumen cluster. Since our system does not rely on a single type of exact segmentation and the object level features are designed to be robust to such limitations then the overall performance of the system is only partially affected. One possible improvement to this system is to adopt a coarse to fine segmentation strategy. For example, the duct segmentation could be evolved using active contours, prior to feature extraction. Another possible improvement would be to segment the fixative material in order to eliminate its effect on the lumen.

The computing platform used in this study is an Intel Core i7 2600K CPU clocked at 3.6Ghz with 8 GB of RAM. The OS is a 64 bit version of Windows 7. The CAD presented was coded using Matlab R2012a, using the image processing and statistics toolboxes. Standard Matlab implementations (from the Statistics toolbox) were used for all classification models. Image analysis execution time is between 30 seconds and a number of minutes, depending mostly on the number of candidate duct contours found in image.

8.2 Decision Support Interface

A pathologist can use the output of the decision tree as a decision support interface. Feature values are trans-

formed into corresponding Z-scores and the tree forms a flowchart which can be used to make multiple assessments of the patient data. Such an *interpretable* representation does not exist for ensemble methods and facilitated the usage of our CAD beyond a single classification result.

9 Conclusion

We have presented a computerized diagnostics system for the detection of early phase pancreatic cancer in a pre-clinical controlled environment. Classification of complex H&E structures is done via supervised learning, making our system highly reliant on ground truth examples. Biopsy samples from the transgenic mice provide such a reliable set.

Our system is modeled in the spirit of a human physician, making supervised decisions based on domain specific knowledge. Our novel duct detection method is generic, making it applicable to various medical imaging tasks, including different malignancies and different degrees of sensitivity.

Quantitative features based on histological primitive segmentation and classification not only provide promising evaluation results using cross validation and independent testing, but also bear *meaning*. Each and every feature is defined based on clinical aspects. There are two main reasons for choosing such a framework. First, choosing features that do not bear clinical significance will likely overfit the data, since such features ultimately rely on assumptions based on the existing dataset and do not take into account the vast domain specific knowledge in the field of pathology. Second, the output decision tree based on object level clinical features provides an *interpretable procedure* for a physician. We believe that a visually interpretable model based only on quantifiable clinical measures will lead to computer aided diagnostics systems which are more likely to be incorporated in a pathologist's workflow.

Future work will explore the integration of additional structures of interest and the design of different features that bear clinical interpretation. In addition, we will investigate methods for extracting structures of interest without human intervention.

Acknowledgement We would like to thank Edith Suss-Toby, head of imaging and microscopy center of the biomedical core facility in the Bruce Rappaport faculty of medicine, for her tremendous contribution of knowledge and expertise in whole slide imaging.

Conflict of interest Leeor Langer, Yoav Binenbaum, Leonid Gugel, Moran Amit, Ziv Gil and Shai Dekel declare that they have no conflict of interest.

References

1. Bamford, P., Lovell, B.: Unsupervised cell nucleus segmentation with active contours. *Signal Processing* **71**(2), 203–213 (1998)
2. Begelrnan, G., Gur, E., Rivlin, E., Rudzsky, M., Zalevsky, Z.: Cell nuclei segmentation using fuzzy logic engine. In: *Image Processing, 2004. ICIP'04. 2004 International Conference on*, vol. 5, pp. 2937–2940. IEEE (2004)
3. Beucher, S., et al.: The watershed transformation applied to image segmentation. *SCANNING MICROSCOPY-SUPPLEMENT* pp. 299–299 (1992)
4. Caruana, R., Niculescu-Mizil, A.: An empirical comparison of supervised learning algorithms. In: *Proceedings of the 23rd international conference on Machine learning*, pp. 161–168. ACM (2006)
5. Collins, M.A., Bednar, F., Zhang, Y., Brisset, J.C., Galbán, S., Galbán, C.J., Rakshit, S., Flannagan, K.S., Adsay, N.V., di Magliano, M.P.: Oncogenic kras is required for both the initiation and maintenance of pancreatic cancer in mice. *The Journal of clinical investigation* **122**(2), 639 (2012)
6. Doi, K.: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics* **31**(4-5), 198–211 (2007)
7. Doyle, S., Feldman, M., Tomaszewski, J., Madabhushi, A.: A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. *Biomedical Engineering, IEEE Transactions on* **59**(5), 1205–1218 (2012)
8. Doyle, S., Hwang, M., Shah, K., Madabhushi, A., Feldman, M., Tomaszewski, J.: Automated grading of prostate cancer using architectural and textural image features. In: *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, pp. 1284–1287. IEEE (2007)
9. Fairchild, M.D.: *Color appearance models*. John Wiley & Sons (2013)
10. Fischer, A.H., Jacobson, K.A., Rose, J., Zeller, R.: Hematoxylin and eosin staining of tissue and cell sections. *Cold Spring Harbor Protocols* **2008**(5), pdb-prot4986 (2008)
11. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. In: *Computational learning theory*, pp. 23–37. Springer (1995)

12. Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B.: Histopathological image analysis: A review. *Biomedical Engineering, IEEE Reviews in* **2**, 147–171 (2009)
13. Gurcan, M.N., Pan, T., Shimada, H., Saltz, J.: Image analysis for neuroblastoma classification: segmentation of cell nuclei. In: *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pp. 4844–4847. IEEE (2006)
14. He, L., Long, L.R., Antani, S., Thoma, G.R.: Histology image analysis for carcinoma detection and grading. *Computer methods and programs in biomedicine* **107**(3), 538–556 (2012)
15. Hidalgo, M.: Pancreatic cancer. *New England Journal of Medicine* **362**(17), 1605–1617 (2010)
16. Hingorani, S.R., Wang, L., Multani, A.S., Combs, C., Deramandt, T.B., Hruban, R.H., Rustgi, A.K., Chang, S., Tuveson, D.A.: $i\zeta$ $trp53r172h\zeta/i\zeta$ and $i\zeta$ $krasg12d\zeta/i\zeta$ cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer cell* **7**(5), 469–483 (2005)
17. Hruban, R.H., Fukushima, N.: Pancreatic adenocarcinoma: update on the surgical pathology of carcinomas of ductal origin and panins. *Modern Pathology* **20**, S61–S70 (2007)
18. Naik, S., Doyle, S., Agner, S., Madabhushi, A., Feldman, M., Tomaszewski, J.: Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In: *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pp. 284–287. IEEE (2008)
19. Olive, K.P., Tuveson, D.A.: The use of targeted mouse models for preclinical testing of novel cancer therapeutics. *Clinical Cancer Research* **12**(18), 5277–5287 (2006)
20. Petushi, S., Garcia, F.U., Haber, M.M., Katsinis, C., Tozeren, A.: Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. *BMC Medical Imaging* **6**(1), 14 (2006)
21. Plissiti, M.E., Nikou, C., Charchanti, A.: Watershed-based segmentation of cell nuclei boundaries in pap smear images. In: *Information Technology and Applications in Biomedicine (ITAB), 2010 10th IEEE International Conference on*, pp. 1–4. IEEE (2010)
22. Rabinovich, A., Agarwal, S., Laris, C., Price, J.H., Belongie, S.: Unsupervised color decomposition of histologically stained tissue samples. (2003)
23. Roth, D.: Learning based programming. In: *Innovations in Machine Learning*, pp. 73–95. Springer (2006)
24. Roullier, V., L  zoray, O., Ta, V.T., Elmoataz, A.: Multi-resolution graph-based analysis of histopathological whole slide images: Application to mitotic cell extraction and visualization. *Computerized Medical Imaging and Graphics* **35**(7), 603–615 (2011)
25. Sabata, B.: Digital pathology imaging-the next frontier in medical imaging. In: *Advanced Computer Science and Information Systems (ICACSIS), 2012 International Conference on*, pp. 1–6. IEEE (2012)
26. Sertel, O., Kong, J., Shimada, H., Catalyurek, U., Saltz, J.H., Gurcan, M.N.: Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development. *Pattern recognition* **42**(6), 1093–1103 (2009)
27. Veta, M., Huisman, A., Viergever, M.A., van Diest, P., Pluim, J.P.: Marker-controlled watershed segmentation of nuclei in h&e stained breast cancer biopsy images. In: *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pp. 618–621. IEEE (2011)
28. Vezhnevets, A., Barinova, O.: Avoiding boosting overfitting by removing confusing samples. In: *Machine Learning: ECML 2007*, pp. 430–441. Springer (2007)