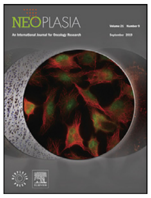
Neoplasia 42 (2023) 100911

Contents lists available at ScienceDirect

Neoplasia

journal homepage: www.elsevier.com/locate/neo



Lung cancer lesion detection in histopathology images using graph-based sparse PCA network☆

Sundaresh Ram[a,*], Wenfei Tang[b], Alexander J. Bell[a], Ravi Pal[c], Cara Spencer[d], Alexander Buschhaus[c], Charles R. Hatt[c,e], Marina Pasca diMagliano[f], Alnawaz Rehemtulla[g], Jeffrey J. Rodríguez[h], Stefanie Galban[c], Craig J. Galban[a]

[a] Departments of Radiology, and Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA
[b] Department of Computer Science and Engineering, University of Michigan, Ann Arbor, MI 48109, USA
[c] Department of Radiology, University of Michigan, Ann Arbor, MI 48109, USA
[d] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA
[e] Imbio LLC, Minneapolis, MN 55405, USA
[f] Departments of Surgery, and Cell and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA
[g] Departments of Radiology, and Radiation Oncology, University of Michigan, Ann Arbor, MI 48109, USA
[h] Departments of Electrical and Computer Engineering, and Biomedical Engineering, The University of Arizona, Tucson, AZ 85721, USA

ARTICLE INFO

Keywords:
Machine learning
Graph-based sparse PCA
Computational imaging
Cancer lesion detection
Image analysis

ABSTRACT

Early detection of lung cancer is critical for improvement of patient survival. To address the clinical need for efficacious treatments, genetically engineered mouse models (GEMM) have become integral in identifying and evaluating the molecular underpinnings of this complex disease that may be exploited as therapeutic targets. Assessment of GEMM tumor burden on histopathological sections performed by manual inspection is both time consuming and prone to subjective bias. Therefore, an interplay of needs and challenges exists for computer-aided diagnostic tools, for accurate and efficient analysis of these histopathology images. In this paper, we propose a simple machine learning approach called the graph-based sparse principal component analysis (GS-PCA) network, for automated detection of cancerous lesions on histological lung slides stained by hematoxylin and eosin (H&E). Our method comprises four steps: 1) cascaded graph-based sparse PCA, 2) PCA binary hashing, 3) block-wise histograms, and 4) support vector machine (SVM) classification. In our proposed architecture, graph-based sparse PCA is employed to learn the filter banks of the multiple stages of a convolutional network. This is followed by PCA hashing and block histograms for indexing and pooling. The meaningful features extracted from this GS-PCA are then fed to an SVM classifier. We evaluate the performance of the proposed algorithm on H&E slides obtained from an inducible *K-ras*^{G12D} lung cancer mouse model using precision/recall rates, F_β-score, Tanimoto coefficient, and area under the curve (AUC) of the receiver operator characteristic (ROC) and show that our algorithm is efficient and provides improved detection accuracy compared to existing algorithms.

Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide, with an estimated 1.6 million deaths each year [1]. Development of novel therapies to battle lung cancer has been greatly aided by the emergence of genetically engineered mouse models (GEMMs) of lung cancer, such as the *K-ras*^{G12D}; *p53*^{Frt} non-small-cell lung carcinoma (NSCLC) model, where the compound effect of conditional mutations in the *K-ras* oncogene and the *p53* tumor suppressor gene leads to development of adenocarcinomas in the mouse lung [2,3]. Since GEMMs recapitulate certain aspects of the human disease associated with the stroma, vascularity, and immune infiltrate better than other models, it is important to be able to detect, identify and localize the lung tumor lesions seen on the histopathological sections as shown in Fig. 1.

Manual assessment of tumor burden (the amount of tumor cells/mass present in a subject's body) on histopathological mouse lung sections is difficult, time consuming, and a labor-intensive process. This is due to various reasons such as fluctuating intensities [4], color change and morphological variations within structures of the cancer lesions in these images [5], tumor heterogeneity [6] (see Fig. 1), low signal-to-noise-

☆ This work was supported in part by the National Heart, Lung, and Blood Institute of the National Institute of Health under Grant R01HL139690 and in part by grant support awarded to Dr. Stefanie Galban by the Rogel Cancer Center at the University of Michigan.

* Corresponding author.
E-mail address: sundarer@umich.edu (S. Ram).

https://doi.org/10.1016/j.neo.2023.100911
Received 7 March 2023; Accepted 17 May 2023

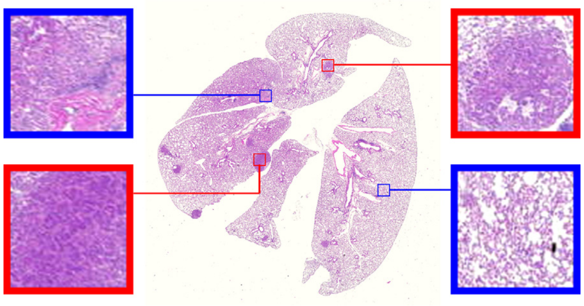


Fig. 1. An example whole-slide histopathological image from our dataset consisting of many tumor lesions. The high-resolution inset images show the visual features that characterize the tumor (red frame) and normal (blue frame) regions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ratio [7,8], variations in illumination [9], microscopy imaging limitations [10–13], and the large number of images and the number of lesions per image an expert has to demarcate. Moreover, the task of manual detection of cancer lesions on H&E slides can be subjective, leading to inter-observer variability. Therefore, there is a pressing need for computer-aided diagnostic tools for accurate and efficient quantitative analysis of histopathology images [14–17].

Tumor detection and classification tools within the commonly available microscopy software are based on feature extraction techniques such as size, shape, and morphological features [8,14–16,18–20], texture features including local binary pattern (LBP) [21–23], local Fourier transform [24], co-occurrence matrix and fractal texture features [25], and energy minimization and optimization-based techniques [26–29]. These techniques suffer considerably due to over-generalization and therefore need extensive customization for the dataset at hand, limiting their use to very simple images obtained/collected in a carefully constrained environment [8]. Tumor detection and grading using size, shape and other morphological features does not work well when the cell population exhibits a variety of sizes and shapes, or when the signal-to-noise (SNR) ratio is poor [30]. Energy minimization and optimization techniques minimize the internal energy within tumor areas for their accurate detection, but may lead to false detections for highly textured and heterogeneous tumor lesions. To overcome these limitations, existing software tools allow user-friendly interfaces to correct the results obtained. This, however, results in losing the benefits of automation such as speed and reproducibility.

There has been much interest in developing algorithmic methods that adapt naturally to the dataset and perform feature discovery. One such popular class of learning or feature discovery methods includes those based on sparse representation-based classification (SRC) [31]. There have been many SRC methods that have been successfully applied to a variety of histopathological image classification problems [32–35]. These methods are based on finding linear representations in the data. However, linear representations are almost always inadequate for representing non-linear structures of the data which arise in many practical applications. A recent class of learning-based methods involve the design of deep neural networks that can be trained to learn relevant features by themselves. There have been plenty of deep learning methods that have been developed for histopathological image classification [5,36–42]. The success of deep learning, however, has been fueled by the availability of generous and clean training data. When the training data is limited and/or noisy, as is often the case in medical imaging, these methods tend to show a performance degradation [43]. Another class of learning-based approaches involve orthogonal transformation of the data such as principal component analysis (PCA) transform to extract relevant features for image classification [30,44–46]. These learning-based approaches using orthogonal transformation explore the data distribution to preserve global structures in the data.

In this paper, we present a simple machine learning approach called the graph-based sparse principal component analysis (GS-PCA) network, which combines the local and global structures of all the data and is implemented in a deep learning framework to learn an explicit nonlinear mapping of the data for accurate detection and classification. We use the most basic and easy operations to emulate the processing stages in a typical (convolutional) neural network: First, graph-based sparse PCA filters are used as the data-adapting convolutional filter bank at each stage of the network. Next, we perform a simple binary quantization (hashing) that serves as the nonlinear stage, followed by block-wise histograms of the binary codes as the feature pooling stage to obtain the final output features of the network. Finally, we train a support vector machine (SVM) classifier on the output features of the network to obtain the final classification instead of the regular softmax classifier, as the softmax classifier is known to overfit [44]. For ease of reference, we call this data-processing network a *Graph-Based Sparse PCA Network* (GS-PCANet). The key contributions of this paper are as follows:

- **Feature Extraction Using Graph-Based Sparse PCA:** Unlike other histopathology image classification methods, in this work we propose a baseline neural network method called GS-PCANet, which is different from prior methods [30,44–46] in two aspects. 1) We include an additional sparsity promoting term in the PCA transformation so as to select more interpretable features from the image patches. 2) We include a graph regularization term in the objective function to recover the low-dimensional manifold structure from high-dimensional sampled data.
- **Computationally Efficient Approach:** Our proposed GS-PCANet is computationally efficient in comparison to other deep learning methods in two aspects. 1) We show that a simple two-stage network is good enough to extract all the relevant features for classifying the tumor versus healthy lung regions. 2) We do not need to learn the filter weights at each stage of the network.

We evaluate the proposed method and seven state-of-the-art algorithms developed for histopathology image classification on a dataset of 67 images provided by the Stefanie Galban Lab, at the University of Michigan. The dataset consists of microscopy images of murine H&E stained lung sections and are divided into two categories: images of non-tumor-bearing control mice and images of mice with visible tumor.

Principal Component Analysis

Let X denote an $n \times p$ matrix of n rows and p columns of rank $q \leq \min(n, p)$, where n is the number of data samples, and p is the number of features/variables in each data sample. Let x_i for $i = 1, \ldots, n$ denote a row of the matrix X, assumed to have a zero mean. Let Σ denote the covariance matrix of x_i, where Σ is a positive definite matrix of size $p \times p$, which can be decomposed as

$$\Sigma = \sum_{i=1}^{q} \sigma_i \mathrm{v}_i \mathrm{v}_i^\top \tag{1}$$

where σ_i is the i^{th} largest eigenvalue of Σ and $\mathrm{v}_i = \left[v_{i1}, \ldots, v_{ip}\right]^\top$ is its associated eigenvector. PCA reduces the dimensionality of the data from p to q by replacing the original features/variables with q linear combinations of the form Xv_k, $k = 1, \ldots, q$ known as the principal components (PCs), which are obtained by maximizing their variance:

$$\mathrm{v}_k = \underset{\mathrm{v}}{\operatorname{argmax}} \left\{ \operatorname{Var}(\mathrm{Xv}) \right\} \quad \text{subject to} \quad \mathrm{v}_k^\top \mathrm{v}_k = 1$$

and

$$\mathrm{v}_j^\top \mathrm{v}_k = 0 \quad \text{for} \quad j < k$$

where v_k is the k^{th} principal loading vector and the projection of the data Xv_k is the k^{th} principal component and the operator $\operatorname{Var}(\cdot)$ denotes the (estimated) variance of a random variable.

Generally, PCA is computed using singular value decomposition (SVD) of X as

$$\mathrm{X} = \mathrm{USV}^{\top} \tag{2}$$

where the columns of $\mathrm{Z} \triangleq \mathrm{US}$ are the PCs, and the columns of V are the corresponding principal loading vectors (also known as basis vectors) [47]. The matrix S is a $q \times q$ diagonal matrix of ordered singular values $s_1 \geq s_2 \geq \ldots \geq s_q > 0$ and the columns of U and V are orthonormal such that $\mathrm{U}^{\top}\mathrm{U} = \mathrm{V}^{\top}\mathrm{V} = \mathrm{I}_q$. If X is low rank, it is possible to significantly reduce its dimensionality by using the q most significant basis vectors. The projection of the data X upon the first q basis vectors gives the PCs.

An alternative formulation for PCA can be derived on the projection framework [44], where the PC loading matrix V also known as the PCA basis (defined as the matrix containing the principal loading vectors) can be estimated by solving the following least squares optimization problem:

$$\min_{\mathrm{A}} \left\{ \| \mathrm{X} - \mathrm{XAA}^{\top} \|_F^2 \right\} \quad \text{subject to} \quad \mathrm{A}^{\top}\mathrm{A} = \mathbb{I}_q \tag{3}$$

where $\| \cdot \|_F$ is the Frobenius norm, $\mathrm{A} \in \mathbb{R}^{p\times q}$ is a matrix whose columns form an orthonormal basis $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_q\}$, and $\mathbb{I}_q$ is an identity matrix of size $q \times q$. The columns of A that minimize (3) are referred to as the PCA basis V. The minimization is solved by formulating it as a least absolute shrinkage and selection operator (LASSO) problem [48]. Each principal component is derived from a linear combination of all p features, consequently making $\boldsymbol{\alpha}$ non-sparse. We use this alternative formulation for PCA feature extraction in this work.

Proposed Method

Based on the PCA methodology, we propose a simple and efficient machine learning method for histopathology image classification. First, we obtain graph-based sparse PCA filters (i.e., the PCs) from the training images as the data adaptive convolutional filter bank for the various stages of a convolutional neural network. Then we perform a simple binary quantization (hashing), which serves as a nonlinear stage. Next, we use block-wise histograms of the binary codes obtained from the quantization process to get the output features of the network. Finally, we train a SVM classifier using the output features to obtain the final classification. The proposed GS-PCANet model is shown in Fig. 2, illustrating each of the above steps involved in our algorithm.

Graph-Based Sparse PCA

From the analysis of PCA in Section II, we can obtain a sparse PCA basis by including a regularization term in (3). Inclusion of a sparsity penalty reduces the number of features involved in each linear combination for obtaining the PCs. One way to extend (3) to obtain sparse basis vectors is by imposing ℓ_1-norm and ℓ_2-norm penalty constraints upon the regression coefficients (basis vectors) [48]:

$$\min_{\mathrm{A,B}} \left\{ \|\mathrm{X} - \mathrm{XBA}^{\top}\|_F^2 + \lambda \sum_{j=1}^{q} \|\boldsymbol{\beta}_j\|_2^2 + \sum_{j=1}^{q} \lambda_{1,j} \|\boldsymbol{\beta}_j\|_1 \right\} \tag{4}$$

$$\text{subject to} \quad \mathrm{A}^{\top}\mathrm{A} = \mathbb{I}_q$$

where the same λ (the regularization parameter of the ℓ_2-norm) is used for all q components, different $\lambda_{1,j}$'s (the regularization parameters of the ℓ_1-norm) are allowed for penalizing the loadings of different PCs. The $\mathrm{B} \in \mathbb{R}^{p\times q}$ corresponds to the required sparse basis $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_q\}$. The ℓ_1-norm and ℓ_2-norm regularization terms penalize the number of non-zero coefficients in $\boldsymbol{\beta}$, whereas the loss term simultaneously minimizes the reconstruction error $\|\mathrm{X} - \mathrm{XBA}^{\top}\|_F^2$. If λ and the $\lambda_{1,j}$'s are zero, the problem reduces to finding the ordinary PCA basis vectors, equivalent to (3). When λ, $\lambda_{1,j}$'s > 0 some coefficients of $\boldsymbol{\beta}_j$ are forced to zero, resulting in sparsity.

The sparse PCA defined in (4) preserves the global structures in the data, but does not retain intrinsic geometric structure within input data, thereby missing mutual influences in the data. In addition to preserving the global structures, we are interested in preserving the local structures, i.e., k nearest neighbor (kNN) preservation of each data sample x_i, as they help in identifying local features in the data (See Section IV-F for more details). The idea of a graph-Laplacian from manifold learning theory is to recover low-dimensional manifold structure from high-dimensional sampled data [49]. This provides a motivation to embed a Laplacian to PCA to help preserve the local features in the data. Let the vertices $1, \ldots, n$ correspond to data samples $\mathrm{x}_1, \ldots, \mathrm{x}_n$, respectively. We define a symmetric weight matrix $\mathrm{E} = [e_{lm}]_{l,m=1}^{n} \in \mathbb{R}^{n\times n}$, where e_{lm} is the weight of the edge connecting vertices l and m. The value of e_{lm} is set as follows:

$$e_{lm} = \begin{cases} 1, & \text{if} \quad \mathrm{x}_l \in N_k(\mathrm{x}_m) \quad \text{or} \quad \mathrm{x}_m \in N_k(\mathrm{x}_l) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where the set $N_k(\mathrm{x}_l)$ is the set of k nearest neighbors of x_l. Let us suppose that $\mathrm{A}^{\top} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_p\} \in \mathbb{R}^{q\times p}$ is the embedding coordinates of the data and define C as the diagonal matrix with $\mathrm{C}_{ll} = \sum_{m=1}^{n} e_{lm}$. Thus, with weight matrix E, we can formulate a graph regularization term as

$$\frac{1}{2}\sum_{l,m=1}^{n} e_{lm} \|\boldsymbol{\alpha}_l - \boldsymbol{\alpha}_m\|_2^2 = \sum_{k=1}^{n} \boldsymbol{\alpha}_l^{\top} e_{mm} \boldsymbol{\alpha}_l - \sum_{l,m=1}^{n} \boldsymbol{\alpha}_l^{\top} e_{lm} \boldsymbol{\alpha}_l$$
$$= \mathrm{Tr}(\mathrm{ACA}^{\top}) - \mathrm{Tr}(\mathrm{AEA}^{\top}) = \mathrm{Tr}(\mathrm{ALA}^{\top}) \tag{6}$$

where L is the graph Laplacian matrix computed as $\mathrm{L} = \mathrm{C} - \mathrm{E}$ and Tr is the trace of a matrix. Simply put, in the case of maintaining the local adjacency relationship of the graph, the graph can be transformed from the high-dimensional space to a low-dimensional space. Minimizing the graph regularization term in (6) helps to preserve the low-dimensional manifold structure in the data [49]. Combining the sparse PCA from

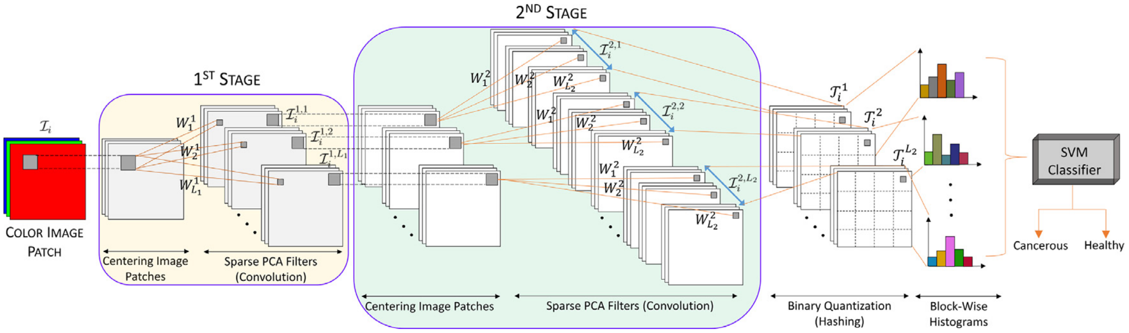


Fig. 2. An outline of the proposed (two-stage) GS-PCANet.

(4) and the graph regularization from (6), we propose a graph-based sparse PCA model,

$$\min_{\mathrm{A,B}} \left\{ \|\mathrm{X} - \mathrm{XBA}^{\top}\|_F^2 + \lambda \sum_{j=1}^{q} \|\boldsymbol{\beta}_j\|_2^2 + \sum_{j=1}^{q} \lambda_{1,j} \|\boldsymbol{\beta}_j\|_1 + \rho\,\mathrm{Tr}(\mathrm{ALA}^{\top}) \right\} \quad (7)$$

subject to $\mathrm{A}^{\top}\mathrm{A} = \mathbb{I}_q$

where ρ is a graph regularization parameter. To solve (7), we perform the following steps: first solve an ordinary PCA problem to fix A, then formulate an elastic net with the fixed A and solve for B, then perform SVD to update $\boldsymbol{\alpha}$, and repeat these steps until convergence, finally obtaining the solution as $\mathrm{V} = \mathrm{B}/\ \|\mathrm{B}\|$.

Architecture of GS-PCA Network

Suppose there are N training images $\{I_i\}_{i=1}^{N}$ of size $u \times v$, and assume that PCA filter size is $t_1 \times t_2$ (formed by reshaping a basis vector of length $t_1 \times t_2$) at all stages of the network. The sparse PCA filters are learned from these training images. We describe each component of the network in detail below (see Fig. 2).

1) First stage (GS-PCA)

For each training image I_i, around each pixel we take an image patch of size $t_1 \times t_2$ and denote all the overlapping image patches in the i^{th} image as $\mathrm{X}_i = \left[\mathrm{x}_{i,1}, \mathrm{x}_{i,2}, \ldots, \mathrm{x}_{i,\tilde{u}\tilde{v}}\right]$, where $\mathrm{x}_{i,j}$ denotes the j^{th} vectorized image patch in I_i, $\tilde{u} = u - \left(t_1 - 1\right)$, $\tilde{v} = v - \left(t_2 - 1\right)$. We then subtract the image patch mean from each of the image patches and obtain the centralized matrix $\bar{\mathrm{X}}_i$ of X_i as $\bar{\mathrm{X}}_i = \left[\bar{\mathrm{x}}_{i,1}, \bar{\mathrm{x}}_{i,2}, \ldots, \bar{\mathrm{x}}_{i,\tilde{u}\tilde{v}}\right]$, where $\bar{\mathrm{x}}_{i,j} = \mathrm{x}_{i,j} - \mu_i$ and $\mu_i = \mathbb{E}\left[\mathrm{x}_i\right] \approx (\frac{1}{n}) \sum_{j=1}^{n} \mathrm{x}_{i,j}$. By constructing a similar centralized matrix for each training image I_i, we obtain

$$\mathrm{X} = \left[\bar{\mathrm{X}}_1, \bar{\mathrm{X}}_2, \ldots, \bar{\mathrm{X}}_N\right] \in \mathbb{R}^{t_1 t_2 \times N\tilde{u}\tilde{v}}. \quad (8)$$

Assuming that we have L_i PCA filters in stage i, sparse PCA minimizes the reconstruction error within a family of orthonormal filters using (7), where $\mathbb{I}_q$ is an identity matrix of size $L_1 \times L_1$. The solution to the minimization problem in (7) are the L_1 principal eigenvectors of $\mathrm{XX}^{\top}$ [44]. The PCA filters can therefore be expressed as

$$\boldsymbol{W}_{l_1}^{1} \doteq \mathrm{Mat}_{t_1,t_2}\left[\mathcal{Q}_{l_1}\left\{\mathrm{XX}^{\top}\right\}\right] \in \mathbb{R}^{t_1 \times t_2}, \quad l_1 = 1, 2, \ldots, L_1 \quad (9)$$

where $\mathrm{Mat}_{t_1,t_2}[d]$ is an operator that reshapes a column vector $d \in \mathbb{R}^{t_1 t_2}$ to a matrix $\boldsymbol{W} \in \mathbb{R}^{t_1 \times t_2}$ and $\mathcal{Q}_{l_1}\left\{\mathrm{XX}^{\top}\right\}$ denotes the l_1^{th} principal eigenvector of $\mathrm{XX}^{\top}$. The L_1 principal eigenvectors capture the main variation of the centralized image patches in the training data. Similar to a convolutional neural network we stack multiple stages of the sparse PCA filters to extract higher level features.

2) Second stage (GS-PCA)

We repeat the same process as in first stage. Let the l^{th} filter output of first stage be

$$I_i^{1,l_1} \doteq I_i * \boldsymbol{W}_{l_1}^{1}, \qquad i = 1, 2, \ldots, N \quad (10)$$

where $*$ denotes 2D convolution and boundary of the images I_i are zero padded before convolution. Similar to the first stage we collect all the overlapping image patches of the convolved image I_i^{1,l_1}, subtract the patch mean from each patch and obtain the centralized matrix $\bar{\mathrm{Y}}_i^{l_1} = \left[\bar{\mathrm{y}}_{i,1}, \bar{\mathrm{y}}_{i,2}, \ldots, \bar{\mathrm{y}}_{i,\tilde{u}\tilde{v}}\right]$, where $\bar{\mathrm{y}}_{i,j}$ is the j^{th} mean subtracted image patch in I_i^{1,l_1}. We define $\mathrm{Y}^{l_1} = [\bar{\mathrm{Y}}_1^{l_1}, \bar{\mathrm{Y}}_2^{l_1}, \ldots \bar{\mathrm{Y}}_N^{l_1}]$ as the matrix containing all the mean subtracted patches of the l^{th} filter output and concatenate Y^{l} for all filter outputs as

$$\mathrm{Y} = \left[\mathrm{Y}^1, \mathrm{Y}^2, \cdots, \mathrm{Y}^{L_1}\right] \in (\mathbb{R})^{t_1 t_2 \times L_1 N\tilde{u}\tilde{v}} \quad (11)$$

Once again we solve (7) with Y as the input. The solution to the minimization problem in (7) are the L_2 principal eigenvectors of $\mathrm{YY}^{\top}$. The sparse PCA filters of the second stage are then obtained as

$$\boldsymbol{W}_{l_2}^{2} \doteq \mathrm{Mat}_{t_1,t_2}\left[\mathcal{Q}_{l_2}\left\{\mathrm{YY}^{\top}\right\}\right] \in \mathbb{R}^{t_1 \times t_2}, \quad l_2 = 1, 2, \ldots, L_2. \quad (12)$$

For each input image I_i^{1,l_1} of the second stage, there will be L_2 output images of size $u \times v$ generated as

$$I_i^{2,l_2} \doteq \left\{ I_i^{1,l_1} * \boldsymbol{W}_{l_2}^{2} \right\}_{l_2=1}^{L_2} \quad (13)$$

After the second stage we will obtain $L_1 L_2$ output images. It is easy to repeat the above process to build more (sparse PCA) stages if a deeper architecture is needed.

3) Binary quantization (hashing)

For each of the L_1 input images I_i^{1,l_1} presented to the second stage we obtain L_2 real-valued output images I_i^{2,l_2}. We binarize these outputs and obtain $\{H(I_i^{1,l_1} * \boldsymbol{W}_{l_2}^{2})\}_{l_2=1}^{L_2}$, where $H(\cdot)$ is a Heaviside step (like) function, which has a value of 1 for positive entries and zero otherwise. Around each pixel, we view the vector of L_2 binary bits as a decimal number, thus converting the L_2 outputs in I_i^{2,l_2} into a single integer-valued "image"

$$T_i^{l_1} \doteq \sum_{l_2=1}^{L_2} 2^{l_2-1} H\left(I_i^{1,l_1} * \boldsymbol{W}_{l_2}^{2}\right), \quad (14)$$

which has pixel values in the range $\left[0, 2^{L_2} - 1\right]$.

4) Block-wise histograms

We partition each of the L_1 "images" $T_i^{l_1}, l_1 = 1, 2, \ldots, L_1$ into G distinct blocks, compute the histogram (with 2^{L_2} bins) of the decimal values in each block and concatenate all G histograms into a single vector denoting it as $G_{\text{hist}}(T_i^{l_1})$. After such an encoding process the "feature" of the input image I_i is then defined to be the set of block-wise histograms, i.e.,

$$f_i \doteq \left[G_{\text{hist}}\left(T_i^{1}\right), \ldots, G_{\text{hist}}\left(T_i^{L_1}\right)\right] \in \mathbb{R}^{\left(2^{L_2}\right) L_1 G}. \quad (15)$$

We use overlapping blocks to build the feature vector for each input image I_i as it helps in retaining most amount of the information.

We train a linear support vector machine (SVM) classifier [50] using the feature vector f_i obtained for each input image I_i from the GS-PCANet in order to classify cancer lesions versus normal tissues on H&E stained histological lung slides.

Classifying Color Images

There are several options to extend the proposed GS-PCANet method to be able to extract features for classifying color images. In this work, we follow the approach described in Gurcan et al. [14], Chan et al. [44] and apply the proposed GS-PCANet to each of the red, blue, and green channels to obtain multichannel sparse PCA filters, that are then used to extract features for classifying the color images.

Experiments and Results

In this section we evaluate our proposed GS-PCANet image classification algorithm with other open-source histopathology image classification methods: SpPCANet method for image classification [46], multiple clustered instance learning (MCIL) for histopathology image classification [51], saliency-based dictionary learning (SDL) [34], analysis-synthesis learning with shared features (ASLF) [35], patch-based convolutional neural network (PCNN) [36], encoded local projections (ELP) for histopathology image classification [20], and weakly supervised deep learning (WSDL) for whole slide tissue classification [40]. We evaluate these seven methods using commonly used detection/classification measures: precision (P), recall (R), detection accuracy, F_β-score, Tanimoto coefficient (T), and the receiver operating characteristic (ROC) curves along with the area under the curve (AUC).

The Precision P and recall R (a.k.a. true positive rate or sensitivity) are given by

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \quad \mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{16}$$

where TP is the number of true positive classifications, FP is the number of false positive classifications, and FN is the number of false negative classifications. The false positive rate (a.k.a. complement of specificity) is defined as FP/(FP + FN). An ROC curve is a plot of the true positive rate versus the false positive rate. The detection accuracy is defined as (TP + TN)/(TP + FP + TN + FN).

The F_β-score is defined by

$$F_\beta = \left(1+\beta^2\right)\frac{\mathrm{P\,R}}{\left(\beta^2\mathrm{P}\right)+\mathrm{R}} \tag{17}$$

We use F_1 (i.e., $\beta = 1$) as this is the most common choice for this type of evaluation [8].

Tanimoto coefficient, also known as Tanimoto distance in statistics, is defined as

$$\mathrm{T} = \frac{\mathrm{TP}}{\mathrm{M}+\mathrm{N}-\mathrm{TP}} \tag{18}$$

where M is the number of detected individual tumors by an automated algorithm and N is the actual number of individual tumors in the image.

The AUC is the average of precision P(R) over the interval ($0 \le \mathrm{R} \le 1$), where P(R) is a function of recall R. It is given by

$$\mathrm{AUC} = \int_0^1 \mathrm{P(R)}\, d\mathrm{R}. \tag{19}$$

The best detection algorithm among several alternatives is commonly defined as the one that maximizes the Tanimoto coefficient, AUC, and the F_β-score.

Dataset

The proposed method was mainly developed with the goal of identifying individual tumors in H&E stained whole slide histopathology lung images obtained from an inducible $K\text{-}ras^{G12D}$ lung cancer model. The images were produced using a digital slide scanner (Super COOLSCAN 5000 ED Digital Slide Scanner; Nikon Corporation) with a 1× objective lens (level-0 pixel size: 0.52 μm × 0.52 μm). In our experiments, the size of each image acquired is approximately 3000×3000 pixels. Our dataset consists of a total of 67 whole slide histopathology lung images obtained from 32 non-tumor-bearing mice and 35 mice with visible tumors. All animals were maintained in accordance with the University of Michigan's Institutional Animal Care and Use Committee guidelines approved protocol (UCUCA PRO00008646). A careful manual delineation of the borders of the individual tumors within the 35 images was performed by an expert and considered as ground truth for subsequent analysis. We divide each image in our dataset into non-overlapping image patches of size 20×20 pixels consisting of a total of 52,487 cancer lesion patches and 1,455,023 normal patches.

Experimental Setup

We used a total of 15 non-tumor-bearing mice images and 15 images with visible tumors for training the compared algorithms, consisting of a total of 21,934 cancer lesion patches and 653,092 normal patches. Our test dataset consists of 17 non-tumor-bearing mice images and 20 images with visible tumors consisting of a total of 30,553 cancer lesion patches and 801,931 normal patches. The hyper-parameters of the GS-PCANet algorithm include the filter size (t_1, t_2), the number of stages, the number of filters in each stage (L_1, L_2), and the block size for the local histograms in the output stage. The optimal values for these parameters were automatically selected on a validation set (randomly chosen from within the training data), using the ROC curves by varying one parameter at a time while keeping the others fixed and choosing that value of the parameter that maximizes the AUC of the ROC curve. The parameters of the GS-PCANet were set to $t_1 = t_2 = 5$, $L_1 = 9$, $L_2 = 9$, and, a histogram block size of 8×8.

Qualitative Results

Fig. 3 shows the qualitative detection results for an example image containing visible tumors from our test dataset. Fig. 3(a) shows that the proposed GS-PCANet method detects most of tumor regions correctly with very few false positives and false negatives. Fig. 3(e) shows that the

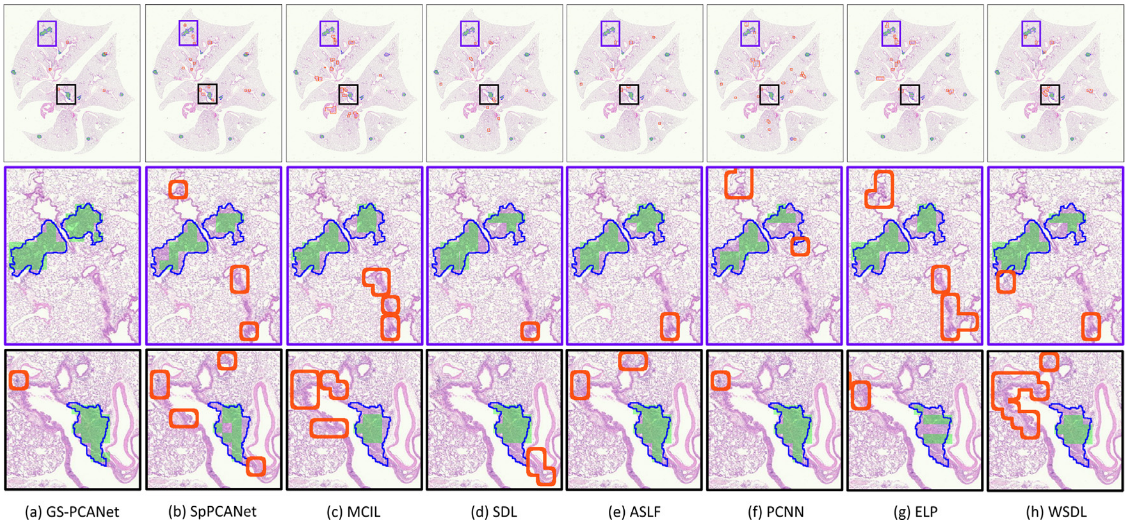


Fig. 3. Detection results on a representative image containing visible tumors in our test dataset using: (a) GS-PCANet, (b) SpPCANet, (c) MCIL, (d) SDL, (e) ASLF, (f) PCNN, (g) ELP, and (h) WSDL. The true borders delineated by an expert of each individual tumor in the image are shown in blue, the true positives patches identified by each method are shown in green and the false positives of each method are bordered in red in the color version of this paper. False negatives are those regions within the blue-bordered individual tumors that are not shaded in green. Results on the entire image are shown in row 1, and two zoomed regions are shown in rows 2 and 3.

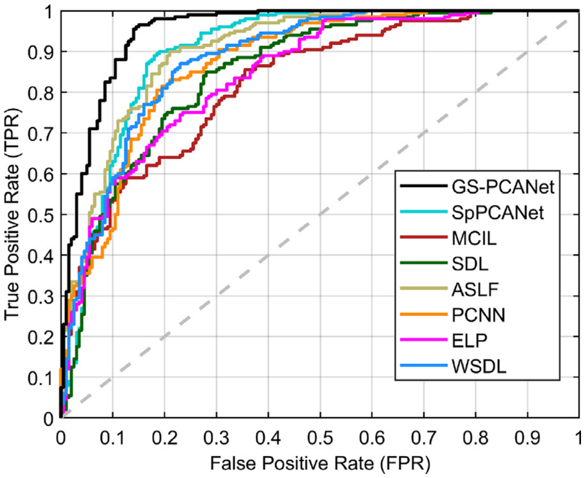


Fig. 4. ROC curve of image patch classification as cancerous or healthy for different methods.

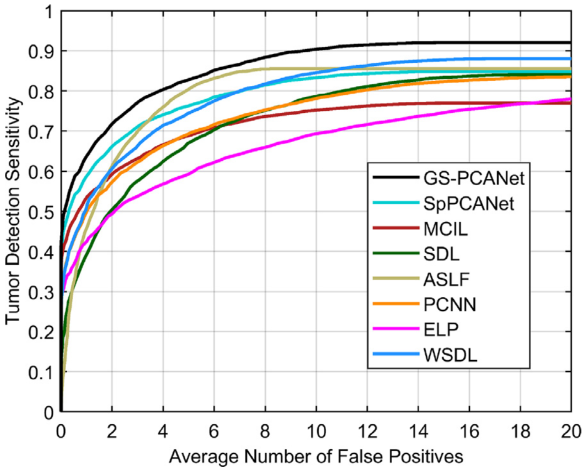


Fig. 5. FROC curve of different methods for the individual tumor detection task within an entire image.

ASLF method is also able to identify the tumor regions well, but detects more false positives than the GS-PCANet method. The SpPCANet, MCIL, and WSDL methods have many misclassifications (with blood vessels being identified as tumors) as shown in Fig. 3(b), (c) and (h), respectively. The ELP method splits a single tumor into three tumors (see Fig. 3(g) row 3), with many false positives. The SDL, PCNN, and ELP methods miss large parts of individual tumors, i.e., have many false negatives as shown in Fig. 3(d), (f), and (g), respectively. Visually it is clear that the proposed GS-PCANet method accurately detects both large and small individual tumors within the whole slide image with very few false positives and false negatives. This is of great significance for those studying oncogenesis, progression, and metastasis because the robustness of the algorithm to the size of the tumor reduces the likelihood that the algorithm will mislabel cases containing only small tumors.

Quantitative Results

We compared the quantitative performance of the automated methods at the image patch level and for the task of individual tumor detection within an entire image as well. Fig. 4 shows the ROC curves of all automated methods at the image patch level on the test dataset. From Fig. 4, we observe that our proposed GS-PCANet method exhibits the most favorable trade-off in terms of accurate detection while maintaining low false positive rate in comparison to the other automated methods. Table 1 shows the quantitative performance of the compared methods for the task of individual tumor detection within the histopathology images in the test dataset. Table 1 shows that the detection accuracy of the proposed GS-PCANet method is much higher than the other competing algorithms. From Table 1, we also observe that the F_β-score, and Tanimoto coefficient (T) of the proposed method are the highest among the compared algorithms. Table 1 also provides the AUC values and their 95% confidence intervals corresponding to the ROC curves in Fig. 4 for each method. We observe from the AUC values that the GS-PCANet method outperforms the alternatives. In addition to the metrics in Table 1, we also computed the free receiver operating characteristics curves (FROC) [8] for all the compared algorithms. Fig. 5 shows that the proposed GS-PCANet method has better tumor detection sensitivity compared to the other automated methods at all points along the FROC curve. This shows that the proposed method detects the individual tumors within these images better than the other compared methods.

The confusion matrix corresponding to competing methods for our test dataset is provided in Table 2. From Table 2, we observe that our proposed GS-PCANet method outperforms competing dictionary learning methods as well as deep learning methods. This could be due to the fact that our proposed GS-PCANet method uses a complete basis representation whereas dictionary learning or deep learning methods use an overcomplete basis to represent the features associated with both normal and cancerous regions within the images.

1) Impact of the number of stages

The impact of the number of GS-PCANet stages for our data is studied here. Specifically, we are interested in the impact on the performance of GS-PCANet when we merge the two stages into one stage that has the equal number of sparse PCA filters and receptive field size defined as the size of the region of the input image that produces the feature. We built a single-stage GS-PCANet (GS-PCANet-1) with $L_1 L_2$ filters of size $(2t_1 - 1) \times (2t_2 - 1)$ and compare it with the two-stage GS-PCANet (GS-PCANet-2) described in Section III-B. The parameters for both networks are set to $t_1 = t_2 = 5$, $L_1 = 9$, $L_2 = 9$, and a histogram block size of 8×8. The detection accuracy and the AUC values of both networks for our test data are reported in Table 3. From Table 3 we observe that the two-stage GS-PCANet outperforms the single-stage GS-PCANet. One explanation behind this could be that in comparison to the PCA filters learned by GS-PCANet-1, the PCA filters of GS-PCANet-2 essentially have a low-

Table 1
Mean performance (and standard deviation) for various algorithms.

Method	Precision (P)	Recall (R)	F_β-score	Tanimoto coefficient (T)	Detection accuracy	AUC
GS-PCANet	**0.872 (0.013)**	**0.955 (0.019)**	**0.912 (0.015)**	**0.903 (0.010)**	**0.908 (0.008)**	**0.951 ± 0.011**
SpPCANet [46]	0.841 (0.019)	0.870 (0.025)	0.855 (0.022)	0.836 (0.014)	0.853 (0.015)	0.907 ± 0.017
MCIL [51]	0.719 (0.022)	0.780 (0.015)	0.748 (0.031)	0.762 (0.019)	0.738 (0.026)	0.821 ± 0.013
SDL [34]	0.752 (0.024)	0.850 (0.031)	0.798 (0.025)	0.801 (0.017)	0.785 (0.011)	0.849 ± 0.021
ASLF [35]	0.811 (0.028)	0.900 (0.019)	0.853 (0.021)	0.829 (0.030)	0.845 (0.018)	0.903 ± 0.022
PCNN [36]	0.807 (0.039)	0.815 (0.031)	0.811 (0.032)	0.796 (0.023)	0.810 (0.024)	0.871 ± 0.039
ELP [20]	0.761 (0.023)	0.750 (0.018)	0.756 (0.021)	0.739 (0.027)	0.758 (0.023)	0.844 ± 0.014
WSDL [40]	0.798 (0.030)	0.785 (0.028)	0.823 (0.031)	0.821 (0.035)	0.818 (0.028)	0.882 ± 0.041

Table 2
Confusion matrix (%).

Class	Cancerous	Healthy	Method
Cancerous	**87.21**	**12.79**	**GS-PCANet**
	84.06	15.94	**SpPCANet** [46]
	71.89	28.11	**MCIL** [51]
	75.22	24.78	**SDL** [34]
	81.08	18.92	**ASLF** [35]
	80.69	19.31	**PCNN** [36]
	76.14	23.86	**ELP** [20]
	79.81	20.19	**WSDL** [40]
Healthy	**4.97**	**95.03**	**GS-PCANet**
	13.47	86.53	**SpPCANet** [46]
	24.04	75.96	**MCIL** [51]
	17.24	82.76	**SDL** [34]
	11.24	88.76	**ASLF** [35]
	18.69	81.31	**PCNN** [36]
	24.63	73.37	**ELP** [20]
	16.04	83.96	**WSDL** [40]

Table 3
Mean performance (and standard deviation) for various stages of GS-PCANet. Here $L_1' = L_1 L_2$ and $t_i' = 2t_i - 1, i = 1, 2$.

Method	Detection accuracy	AUC
GS-PCANet-1 ($L_1' = 81, t_1' = t_2' = 9$)	0.815 (0.019)	0.842 ± 0.014
GS-PCANet-2	**0.908 (0.008)**	**0.951 ± 0.011**

rank factorization, resulting in a lower chance of over-fitting the data. Also, from a computational perspective, GS-PCANet-1 requires learning filters with $L_1 L_2(2t_1 - 1)(2t_2 - 1)$ variables, whereas GS-PCANet-2 only learns filters with a total of $(L_1 + L_2)t_1 t_2$ variables, confirming the need for a multiple-stage network structure. Another benefit of GS-PCANet-2 is the larger receptive field, which leads to observing image regions with a bigger context around the objects of interest, as well as its learning invariance [44] which can capture more semantic information.

Statistical Analysis

To investigate the robustness of training or selection bias for each automated method, we obtain the detection performance for 10 different choices of training image patches (the number of training images were fixed), using the rest of the image patches as test image patches. The detection accuracy for each training run was fit to a Gaussian probability density function (pdf) and plotted in Fig. 6. From Fig. 6, we observe that the mean our proposed GS-PCANet curve is much higher than the competing methods indicating superior average detection accuracy. Even more crucial is the spread/variance of our GS-PCANet curve is smaller than its alternatives indicating highly desirable robustness to the particular choice of training image patches.

We also performed a balanced two-way analysis of variance (ANOVA) [52] on the detection accuracies in the selection-bias experiment for all the methods. Fig. 7 shows these comparisons using a post-hoc Tukey range test [52]. Fig. 7 shows that the performance of the GS-PCANet method is significantly separated from its competing alternatives. p-values of the proposed GS-PCANet method compared with other state-of-the-art methods are observed to be much less than 1×10^{-5}, emphasizing the fact that the GS-PCANet method is more effective.

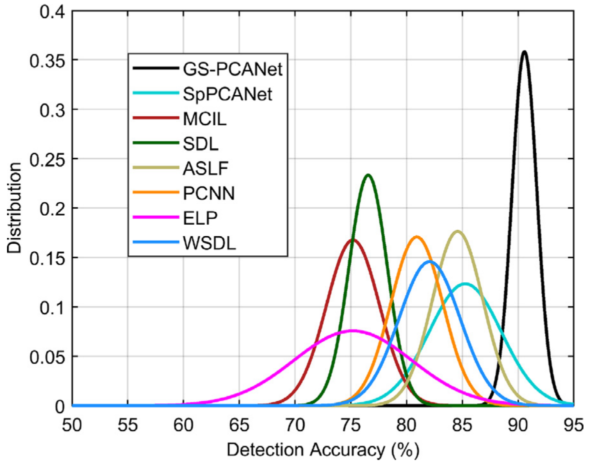


Fig. 6. Selection bias plot showing the distribution of detection accuracy over ten different training choices of image patches for the compared methods.

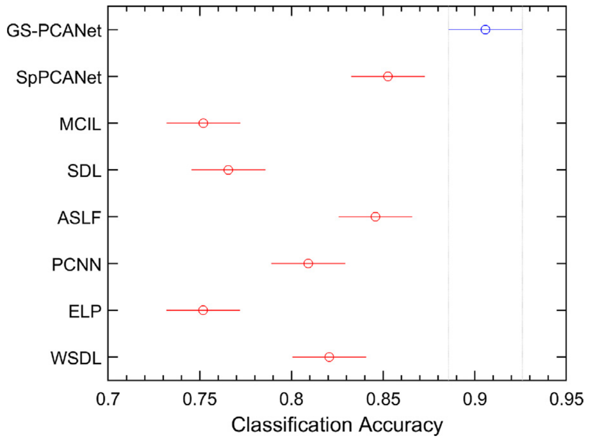


Fig. 7. Comparison of the proposed GS-PCANet method and other state-of-the-art alternatives by a two-way ANOVA. Values reported by ANOVA (using MATLAB function *anova2*) across the methods are $SS = 0.2103, df = 7, MS = 0.0300, F = 36.27, p \ll 1e-5$, indicating that the improved accuracy of the proposed GS-PCANet method is statistically significant. The intervals shown represent 95% confidence intervals of the detection accuracies for the proposed method (blue) and the competing methods (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Comparison of PCA, Sparse PCA, and GS-PCA

It has been shown in the literature that performing a PCA or a sparse PCA analysis preserves the global structures in the data [53], whereas manifold learning-based feature extraction methods are effective for dealing with high-dimensional data as they preserve the local structures in the data via manifold learning [54]. In our GS-PCANet method, we find it reasonable to combine both types of structure-preserving approaches as they strengthen the performance of the image classification task due to providing complementary information. For example, the global structure preservation can improve generalization ability. In this section, we show results of constructing a neural network architecture using the PCA, sparse PCA, and the proposed GS-PCA method. The comparison of the top nine PCs (a.k.a. the filters) of the final stage of the network and the covariance matrix of the PCs for each method are shown in Fig. 8. Different colors in Fig. 8(a)–(c) represent negative (blue), positive (red), and zero-valued (white) coefficients. From Fig. 8(a) to (c), we observe that the GS-PCANet method has more sparse filters as compared to SpPCANet [46] and PCANet [44] methods. Looking at the covariance

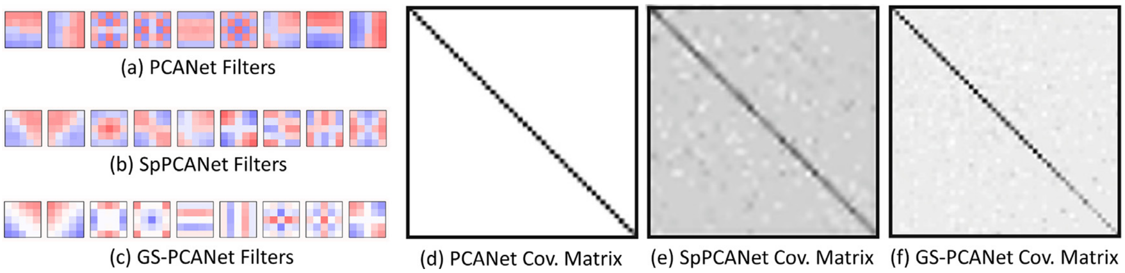
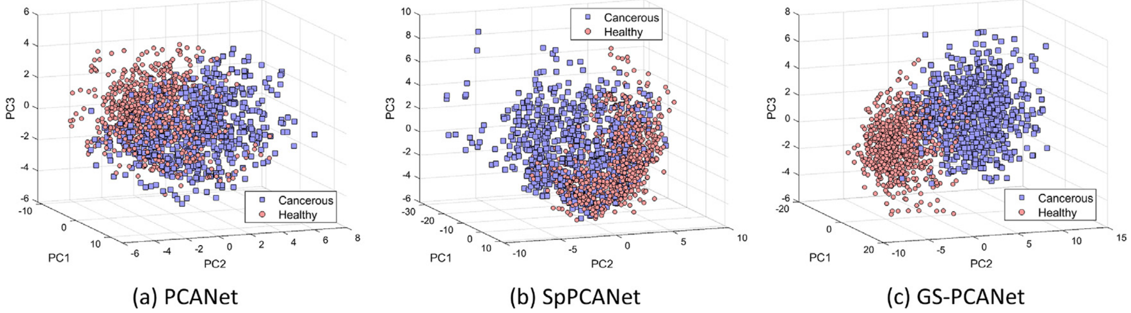

(a) PCANet Filters
(b) SpPCANet Filters
(c) GS-PCANet Filters
(d) PCANet Cov. Matrix
(e) SpPCANet Cov. Matrix
(f) GS-PCANet Cov. Matrix

Fig. 8. Comparison of the filters of the final stage learned on our dataset. Red color are positive values, blue color are negative values, and white color is zero. (a) PCANet filters, (b) SpPCANet filters, (c) GS-PCANet filters. Also shown are the covariance matrix of the components. (d) PCANet covariance matrix, (e) SpPCANet covariance matrix, (f) GS-PCANet covariance matrix.



(a) PCANet
(b) SpPCANet
(c) GS-PCANet

Fig. 9. Scatter plots of the test image patches of our dataset based on the first three principal components by (a) PCANet method, (b) SpPCANet method, (c) GS-PCANet method.

Table 4
Performance metrics for various PCA methods.

Method	Silhouette score	C–H Index	D–B index
PCANet	0.56	185.62	1.04
SpPCANet	0.62	318.24	0.71
GS-PCANet	**0.74**	**362.45**	**0.54**

matrices in Fig. 8(d) to (f), we observe that the PCs for the PCANet are most orthogonal, and that the GS-PCANet method has PCs more orthogonal than those found by SpPCANet method. Additionally, for comparison we present the scatter plots of the top three PCs for each method on image patches from our test dataset in Fig. 9. From Fig. 9, we observe that the GS-PCANet method achieves better separation for the healthy versus the cancerous test image patches in comparison to SpPCANet and PCANet methods. In addition, we computed the mean silhouette score, the Calinski–Harabasz (C–H) index [55], and Davies–Bouldin (D–B) index [56] on these scatter plots and report these values in Table 4. The silhouette score is calculated using the mean intra-cluster distance and the mean nearest cluster distance for each data point. The C–H index (also known as the variance ratio criterion) is defined as the ratio between the within-cluster dispersion and the between-cluster dispersion. The D–B index is defined as the average similarity measure of each cluster with its most similar cluster. The method that has the highest mean silhouette score and C–H index and the lowest D–B index would have the best separation for healthy and cancerous test image patches. From Table 4, we observe that the GS-PCANet method has the highest mean silhouette score and C–H index, and the lowest D–B index among the three methods. These results show that addition of the graph regularization term in the GS-PCANet method leads to a better separation between the image classes in comparison to SpPCANet and PCANet methods.

Table 5
Mean run time (and standard deviation).

Method	Training time (HH:MM:SS)	Run time (Std. Dev.) in Sec.
GS-PCANet	00:21:09	**11.14 (3.09)**
SpPCANet [46]	**00:20:53**	15.21 (1.41)
MCIL [51]	18:25:06	66.35 (14.36)
SDL [34]	01:22:41	46.11 (4.51)
ASLF [35]	01:49:27	19.39 (5.15)
PCNN [36]	19:27:55	39.47 (15.22)
ELP [20]	04:38:03	71.44 (9.40)
WSDL [40]	21:44:17	10.31 (6.02)

Computational Complexity

Here we show computational complexity of the GS-PCANet method by considering a two stage network. For each stage in the GS-PCANet, forming the mean subtracted image patch matrix X has a computational complexity of $\mathcal{O}(t_1 t_2 \tilde{u}\tilde{v})$; the inner product $XX^\top$ in (9) has a complexity of $\mathcal{O}\left(\{t_1 t_2\}^2 \tilde{u}\tilde{v}\right)$; the computational complexity of the eigen decomposition with graph-regularization is $\mathcal{O}\left(\{t_1 t_2\}^3\right)$. The sparse PCA filter convolution has a complexity of $\mathcal{O}(L_i t_1 t_2 uv)$ at stage i. The block-wise histogram computation has a complexity of $\mathcal{O}(uvGL_2)$. With $\tilde{u} = u - (t_1 - 1)$, $\tilde{v} = v - (t_2 - 1)$, and assuming $uv \gg \max\,(t_1, t_2, L_1, L_2, G)$, the overall complexity of GS-PCANet is

$$\mathcal{O}\left(uvt_1t_2\{L_1 + L_2\} + uv\{t_1t_2\}^2\right). \tag{20}$$

The computational complexity in (20) applies to both the training and testing phase of GS-PCANet because the extra computation burden during training is the eigen decomposition, which can be ignored when $uv \gg \max\,(t_1, t_2, L_1, L_2, G)$.

We compared the mean inference run time, namely, the time required to classify all the image patches in a single test image for each of the competing algorithms. Table 5 shows the mean and standard deviation of the run time each method takes to classify an entire image. From Table 5, we observe that the proposed GS-PCANet method runs 0.83 seconds slower than the WSDL method, but is on average faster than all the other methods. The SDL and ASLF methods classify the test image patch by reconstructing them from the learned dictionaries and thus take

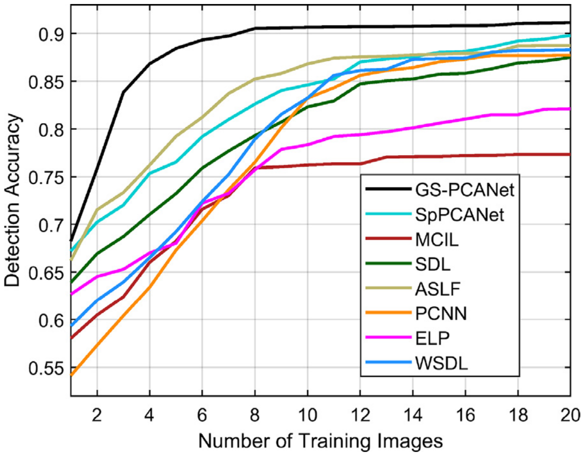


Fig. 10. Detection accuracy as a function of the number of training images for the competing methods.

more time to execute at test time. The ELP algorithm finds the Radon transformation of each test image patch at various orientations, thereby taking more time to classify each test image patch. The MCIL method integrates the clustering of multiple subtypes of a single class into the MIL classification framework, thus requiring more run time compared to the other methods, except for the ELP method. In Table 5 we also report the training time required to train each of the competing algorithms. From Table 5, we observe that the proposed GS-PCANet method and the SpPCANet method take roughly about 21 min to train, whereas the other methods take about 3 to 62 times more time to train a good model. The small training time of the GS-PCANet method is attributed to the low computational complexity of the method.

Impact on Number of Training Images

In this section, we show the practicality and applicability of the proposed GS-PCANet method in medical imaging tasks where we have very few data to learn from. Whereas in all other experiments we trained on 15 images from each class, in this experiment we varied the number of training images (from 1 to 20) for all the competing methods and computed detection accuracy of these methods. Fig. 10 shows the detection accuracy of all the competing algorithms on the test dataset of 27 images (12 non-tumor images and 15 images with visible tumors). From Fig. 10, we observe that the proposed GS-PCANet method trained with as few as 8 images achieves a high detection accuracy of 91%, whereas the other methods are able to achieve a maximum detection accuracy of only about 89% and also require as much as 20 training images. This shows that the proposed GS-PCANet method can produce a good model for image classification with less training data.

Discussion and Conclusion

Tumor burden in histopathological sections is difficult to assess by manual evaluation, as well as by prior automated tumor detection algorithms. To solve this problem, our proposed machine learning algorithm uses a cascaded graph-based sparse PCA transform followed by PCA binary hashing and block-wise histograms to obtain features within image patches. These features are then used to classify an image patch as cancerous or healthy using a linear SVM classifier. Our approach differs from earlier learning-based methods based on deep learning [36,40], instance learning [20,51] or dictionary learning [34,35] for histopathology image classification. Like many deep learning methods, the network parameters, such as the number of stages, the filter size, and the number of filters, need to be optimized and fixed for our GS-PCANet method. Once these parameters are fixed, training the GS-PCANet is extremely simple and efficient because the filter learning in GS-PCANet does not require regularized parameters or require numerical optimization solvers. Moreover, the GS-PCANet consists of only linear operations at each stage with a non-linearity applied only at the output stage, which makes the method more interpretable than other deep learning methodologies.

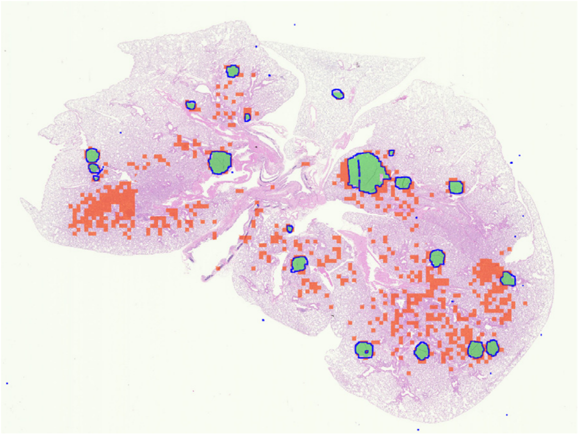


Fig. 11. Example of detection errors produced by all algorithms on an image with visible tumors. The true borders delineated by an expert of each individual tumor in the image are shown in blue, the true positive and the false positive image patches are shown in green and red, respectively, in the color version of this paper (the image is better viewed in zoomed mode). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The GS-PCANet method was first validated with respect to detection accuracy using ROC curves and the AUC of the ROC curve. Second, the algorithm was validated with respect to detection accuracy using the precision, recall, F_β-score, Tanimoto coefficient, FROC curves, and the confusion matrix. Tables 1 and 2 show that the proposed GS-PCANet method performs the best among the compared methods for histopathology image classification. Fig. 3 shows that the proposed GS-PCANet method qualitatively performs the best in comparison to the other methods. Further, Fig. 6 shows that the GS-PCANet method has superior average detection accuracy and is more robust to the choice of training images compared to the other methods. We also show the low computational complexity of the GS-PCANet method and compare the training and inference run times for all the methods. Table 4 shows that the GS-PCANet method is relatively very fast to learn a good model in comparison to other methods. Finally, Fig. 10 shows that the proposed method requires less data to learn a good model.

Next, we present some inherent limitations of the automated methods for tumor detection. Fig. 11 shows an example case of an image containing individual tumors where all algorithms including our algorithm fail to produce optimum detection results. In Fig. 11 we observe that even though the algorithm has detected all the individual tumors, i.e., the true positive image patches shown in green color, it has also detected many false positive image patches shown in red color. On close examination, we see that the false positive image patches within the image look very similar to cancerous image patches. This could be due to the fact that there is not enough resolution in this image to differentiate between the cancerous and healthy image patches, or this histopathology section was captured when some of the underlying cells were transitioning from healthy to being cancerous.

The proposed detection algorithm uses all the image patches in the training data for obtaining the local structures within the data when computing the graph-based term in (6) and (7). This adds to the time complexity and results in noise and outlier image patches still being included. However, the algorithm can be modified by linearly clustering the image patches into subgroups and taking these cluster centers to compute the graph regularization term in (7). Making this change could further reduce detection errors and also accelerate the algorithm, making it more accurate and efficient at the same time.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Dr. Charles Hatt is employed by and has stock options in Imbio, Inc. Dr. Craig Galban is co-inventor of Parametric Response Mapping, which the University of Michigan has licensed to Imbio, Inc., and has a financial interest in Imbio, Inc.

CRediT authorship contribution statement

Sundaresh Ram: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Wenfei Tang:** Formal analysis, Methodology, Software, Validation, Writing – review & editing. **Alexander J. Bell:** Data curation, Project administration, Validation, Visualization, Writing – review & editing. **Ravi Pal:** Data curation, Validation, Visualization, Writing – review & editing. **Cara Spencer:** Data curation, Writing – review & editing. **Alexander Buschhaus:** Data curation, Writing – review & editing. **Charles R. Hatt:** Formal analysis, Investigation, Project administration, Supervision, Writing – review & editing. **Marina Pasca diMagliano:** Resources, Supervision, Writing – review & editing. **Alnawaz Rehemtulla:** Project administration, Supervision, Writing – review & editing. **Jeffrey J. Rodríguez:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Stefanie Galban:** Data curation, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Craig J. Galban:** Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Validation, Visualization, Writing – review & editing.

References

[1] C.S.D. Cruz, L.T. Tanoue, R.A. Matthay, Lung cancer: epidemiology, etiology, and prevention, Clin. Chest Med. 32 (4) (2011) 605–644.

[2] J.C. Walrath, J.J. Hawes, T. Van Dyke, K.M. Reilly, Genetically engineered mouse models in cancer research, in: Advances in Cancer Research, vol. 106, Academic Press, 2010, pp. 113–164.

[3] K.H. Barck, H. Bou-Reslan, U. Rastogi, T. Sakhuja, J.E. Long, R. Molina, A. Lima, P. Hamilton, M.R. Junttila, L. Johnson, R.A.D. Carano, Quantification of tumor burden in a genetically engineered mouse model of lung cancer by micro-CT and automated analysis, Transl. Oncol. 8 (2) (2015) 126–135.

[4] S. Ram, J.J. Rodriguez, Symmetry-based detection of nuclei in microscopy images, in: 2013 IEEE Intl. Conf. Acoustics Speech Signal Process. (ICASSP), IEEE, 2013, pp. 1128–1132.

[5] H. Lin, H. Chen, S. Graham, Q. Dou, N. Rajpoot, P.-A. Heng, Fast scannet: fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection, IEEE Trans. Med. Imaging 38 (8) (2019) 1948–1958.

[6] M.R. Junttila, F.J. de Sauvage, Influence of tumour micro-environment heterogeneity on therapeutic response, Nature 501 (7467) (2013) 346–354.

[7] S. Ram, J.J. Rodriguez, G. Bosco, Segmentation and classification of 3-D spots in FISH images, in: 2010 IEEE Southwest Symp. Image Anal. Interp. (SSIAI), IEEE, 2010, pp. 101–104.

[8] S. Ram, J.J. Rodriguez, Size-invariant detection of cell nuclei in microscopy images, IEEE Trans. Med. Imaging 35 (7) (2016) 1753–1764.

[9] S. Ram, F. Danford, S. Howerton, J.J. Rodriguez, J.P.V. Geest, Three-dimensional segmentation of the ex-vivo anterior lamina cribrosa from second-harmonic imaging microscopy, IEEE Trans. Biomed. Eng. 65 (7) (2018) 1617–1629.

[10] S. Ram, J.J. Rodriguez, G. Bosco, Size-invariant cell nucleus segmentation in 3-D microscopy, in: 2012 IEEE Southwest Symp. Image Anal. Interp. (SSIAI), IEEE, 2012, pp. 37–40.

[11] S. Ram, Sparse Representations and Nonlinear Image Processing for Inverse Imaging Solutions, Department of Electrical and Computer Engineering, The University of Arizona, Tucson, AZ, 2017 Ph.D. dissertation.

[12] S. Ram, M.S. Majdi, J.J. Rodriguez, Y. Gao, H.L. Brooks, Classification of primary cilia in microscopy images using convolutional neural random forests, in: 2018 IEEE Southwest Symp. Image Anal. Interp. (SSIAI), IEEE, 2018, pp. 89–92.

[13] S. Ram, V.T. Nguyen, K.H. Limesand, J.J. Rodriguez, G. Bosco, Combined detection and segmentation of cell nuclei in microscopy images using deep learning, in: 2020 IEEE Southwest Symp. Image Anal. Interp. (SSIAI), IEEE, 2010, pp. 26–29.

[14] M.N. Gurcan, L.E. Boucheron, A. Can, A. Madabhushi, N.M. Rajpoot, B. Yener, Histopathological image analysis: a review, IEEE Rev. Biomed. Eng. 2 (2009) 147–171.

[15] M. Veta, J.P.W. Pluim, P.J. van Diest, M.A. Viergever, Breast cancer histopathology image analysis: a review, IEEE Trans. Biomed. Eng. 61 (5) (2014) 1400–1411.

[16] F. Xing, L. Yang, Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review, IEEE Rev. Biomed. Eng. 9 (2016) 234–263.

[17] S. Ram, W. Tang, A.J. Bell, C. Spenser, A. Buschhuas, C.R. Hatt, M.P. di Magliano, S. Galban, C. Galban, Detection of cancer lesions in histopathological lung images using a sparse PCA network, in: Proc. AACR Virtual Spl. Conf. Artif. Intell. Diag. and Imag., AACR, 2021. Clin Cancer Res 2021; 27(5–Suppl): Abstract nr PO–086

[18] A. Basavanhally, S. Ganesan, M. Feldman, N. Shih, C. Mies, J. Tomaszewski, A. Madabhushi, Multi-field-of-view framework for distinguishing tumor grade in ER+ breast cancer from entire histopathology slides, IEEE Trans. Biomed. Eng. 60 (8) (2013) 2089–2099.

[19] L. Gorelick, O. Veksler, M. Gaed, J.A. Gömez, M. Moussa, G. Bauman, A. Fenster, A.D. Ward, Prostate histopathology: learning tissue component histograms for cancer detection and classification, IEEE Trans. Med. Imaging 32 (10) (2013) 1804–1818.

[20] H.R. Tizhoosh, Representing medical images with encoded local projections, IEEE Trans. Biomed. Eng. 65 (10) (2018) 2267–2277.

[21] S. Reis, P. Gazinska, J.H. Hipwell, T. Mertzanidou, K. Naidoo, N. Williams, S. Pinder, D.J. Hawkes, Automated classification of breast cancer stroma maturity from histological images, IEEE Trans. Biomed. Eng. 64 (10) (2017) 2344–2352.

[22] S. Wan, H.-C. Lee, X. Huang, T. Xu, T. Xu, X. Zeng, Z. Zhang, Y. Sheikine, J.L. Connolly, J.G. Fujimoto, C. Zhou, Integrated local binary pattern texture features for classification of breast tissue imaged by optical coherence microscopy, Med. Imaging Anal. 38 (2017) 104–116.

[23] O. Simon, R. Yacoub, S. Jain, J.E. Tomaszewski, P. Sarder, Multi-radial LBP features as a tool for rapid glomerular detection and assessment in whole slide histopathology images, Sci. Rep. 8 (1) (2018) 1–11.

[24] H. Kong, M. Gurcan, K. Belkacem-Boussaid, Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting, IEEE Trans. Med. Imaging 30 (9) (2011) 1661–1677.

[25] S. Alinsaif, L. Jochen, Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting, BMC Med. Inform. Decis. Mak. 20 (14) (2020) 1–19.

[26] A.B. Tosun, C. Gunduz-Demir, Graph run-length matrices for histopathological image segmentation, IEEE Trans. Med. Imaging 30 (3) (2011) 721–732.

[27] E. Ozdemir, C. Gunduz-Demir, A hybrid classification model for digital pathology using structural and statistical pattern recognition, IEEE Trans. Med. Imaging 32 (2) (2013) 474–483.

[28] B.E. Bejnordi, M. Balkenhol, G. Litjens, R. Holland, P. Bult, N. Karssemeijer, J.A.W.M. van der Laak, Automated detection of DCIS in whole-slide H&E stained breast histopathology images, IEEE Trans. Med. Imaging 35 (9) (2016) 2141–2150.

[29] S. Javed, A. Mahmood, N. Werghi, K. Benes, N. Rajpoot, Multiplex cellular communities in multi-gigapixel colorectal cancer histology images for tissue phenotyping, IEEE Trans. Image Process. 29 (2020) 9204–9219.

[30] J. Shi, J. Wu, Y. Li, Q. Zhang, S. Ying, Histopathological image classification with color pattern random binary hashing-based PCANet and matrix-form classifier, IEEE J. Biomed. Health Inform. 21 (5) (2017) 1327–1337.

[31] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.

[32] U. Srinivas, H.S. Mousavi, V. Monga, A. Hattel, B. Jayarao, Simultaneous sparsity model for histopathological image representation and classification, IEEE Trans. Med. Imaging 33 (5) (2014) 1163–1179.

[33] T.H. Vu, H.S. Mousavi, V. Monga, G. Rao, U.K.A. Rao, Histopathological image classification using discriminative feature-oriented dictionary learning, IEEE Trans. Med. Imaging 35 (3) (2016) 738–751.

[34] R. Sarkar, S.T. Acton, SDL: saliency-based dictionary learning framework for image similarity, IEEE Trans. Image Process. 27 (2) (2018) 749–763.

[35] X. Li, V. Monga, U.K.A. Rao, Analysis-synthesis learning with shared features: algorithms for histology image classification, IEEE Trans. Biomed. Eng. 67 (4) (2020) 1061–1073.

[36] L. Hou, D. Samaras, T.M. Kurc, Y. Gao, J.E. Davis, J.H. Saltz, Patch-based convolutional neural network for whole slide tissue image classification, in: 2016 IEEE Conf. Computer Vis. Pattern Recognit. (CVPR), IEEE, 2016, pp. 2424–2433.

[37] Y. Xu, Z. Jia, L.-B. Wang, Y. Ai, F. Zhang, M. Lai, E.I.-C. Chang, Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features, BMC Bioinform. 18 (1) (2017) 1–17.

[38] D. Tellez, M. Balkenhol, I. Otte-Holler, R. van de Loo, R. Vogels, P. Bult, C. Wauters, W. Vreuls, S. Mol, N. Karssemeijer, G. Litjens, J. van der Laak, F. Ciompi, Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks, IEEE Trans. Med. Imaging 37 (9) (2018) 2126–2136.

[39] F. Xing, T.C. Cornish, T. Bennett, D. Ghosh, L. Yang, Pixel-to-pixel learning with weak supervision for single-stage nucleus recognition in KI67 images, IEEE Trans. Biomed. Eng. 66 (11) (2019) 3088–3097.

[40] G. Campanella, M.G. Hanna, L. Geneslaw, A. Miraflor, V.W.K. Silva, K.J. Busam, E. Brogi, V.E. Reuter, D.S. Klimstra, T.J. Fuchs, Clinical-grade computational pathology using weakly supervised deep learning on whole slide images, Nat. Med. 25 (8) (2019) 1301–1309.

[41] J.W. Wei, L.J. Tafe, Y.A. Linnik, L.J. Vaickus, N. Tomita, S. Hassanpour, Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks, Sci. Rep. 9 (1) (2019) 1–8.

[42] M. Valkonen, J. Isola, O. Ylinen, V. Muhonen, A. Saxlin, T. Tolonen, M. Nykter, P. Ruusuvuori, Cytokeratin-supervised deep learning for automatic recognition of epithelial cells in breast cancers stained for ER, PR, and ki-67, IEEE Trans. Med. Imaging 39 (2) (2020) 534–542.

[43] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016. http://www.deeplearningbook.org

[44] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, Y. Ma, PCANet: a simple deep learning baseline for image classification? IEEE Trans. Image Process. 24 (12) (2015) 5017–5032.

[45] J. Bruna, S. Mallat, Invariant scattering convolution networks, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1872–1886.

[46] K. Dutta, D. Bhattacharjee, M. Nasipuri, SpPCANet: a simple deep learning-based feature extraction approach for 3D face recognition, Multimed. Tools Appl. 79 (41) (2020) 31329–31352.

[47] S.R.S.P. Malladi, S. Ram, J.J. Rodriguez, Image denoising using superpixel-based PCA, IEEE Trans. Multimed. 23 (2020) 2297–2309.

[48] H. Zou, T. Hastie, R. Tibshirani, Sparse principal component analysis, J. Comput. Graph. Stat. 15 (2) (2006) 265–286.

[49] F.R. Chung, Spectral Graph Theory, vol. 92, American Mathematical Soc., 1997.

[50] C. Cortes, V. Vapnik, Support vector networks, Mach. Learn. 20 (3) (1995) 273–297.

[51] Y. Xu, J.-Y. Zhu, E.I.-C. Chang, M. Lai, Z. Tu, Weakly supervised histopathology cancer image segmentation and classification, Med. Imaging Anal. 18 (3) (2014) 591–604.

[52] R.V. Hogg, J. Ledolter, Engineering Statistics, Macmillan Publishing Company, 1987.

[53] X. Zhu, L. Zhang, Z. Huang, A sparse embedding and least variance encoding approach to hashing, IEEE Trans. Image Process. 23 (9) (2014) 3737–3750.

[54] X. Zhu, X. Li, S. Zhang, C. Ju, X. Wu, Robust joint graph sparse coding for unsupervised spectral feature selection, IEEE Trans. Neural Netw. 28 (6) (2017) 1263–1275.

[55] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Commun. Stat. 3 (1) (1974) 1–27.

[56] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1 (2) (1979) 224–227.