Microscopic medical image classification framework via deep learning and shearlet transform

Hadi Rezaeilouyeh
Ali Mollahosseini
Mohammad H. Mahoor

SPIE.

Microscopic medical image classification framework via deep learning and shearlet transform

Hadi Rezaeilouyeh,* **Ali Mollahosseini, and Mohammad H. Mahoor**
University of Denver, Department of Electrical and Computer Engineering, 2155 East Wesley Avenue, Denver, Colorado 80208, United States

Abstract. Cancer is the second leading cause of death in US after cardiovascular disease. Image-based computer-aided diagnosis can assist physicians to efficiently diagnose cancers in early stages. Existing computer-aided algorithms use hand-crafted features such as wavelet coefficients, co-occurrence matrix features, and recently, histogram of shearlet coefficients for classification of cancerous tissues and cells in images. These hand-crafted features often lack generalizability since every cancerous tissue and cell has a specific texture, structure, and shape. An alternative approach is to use convolutional neural networks (CNNs) to learn the most appropriate feature abstractions directly from the data and handle the limitations of hand-crafted features. A framework for breast cancer detection and prostate Gleason grading using CNN trained on images along with the magnitude and phase of shearlet coefficients is presented. Particularly, we apply shearlet transform on images and extract the magnitude and phase of shearlet coefficients. Then we feed shearlet features along with the original images to our CNN consisting of multiple layers of convolution, max pooling, and fully connected layers. Our experiments show that using the magnitude and phase of shearlet coefficients as extra information to the network can improve the accuracy of detection and generalize better compared to the state-of-the-art methods that rely on hand-crafted features. This study expands the application of deep neural networks into the field of medical image analysis, which is a difficult domain considering the limited medical data available for such analysis. © 2016 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.3.4.044501]

1 Introduction

It is estimated that about 1,685,210 new cases of cancer will occur in US during 2016.[1] Prostate and breast cancers are the second most dominant causes of death among cancers in males and females, respectively.[1] Cancer cells can transfer to other tissues and develop new tumors. Hence, it is vital to diagnose and grade cancer in early stages and provide necessary treatment. Histological analysis of tissue slides stained using hematoxylin and eosin (H&E) is the main approach for cancer detection and grading. This process involves a pathologist examining large areas of benign tissue to finally detect the areas of malignancy. Therefore, cancer detection is very time-consuming and challenging. The Gleason grading system is the main approach for prostate grading,[2] which classifies the prostate cancer as grades 1 to 5 with increasing malignancy as the grade increases. The Gleason score is calculated using the sum of the two most dominant Gleason grades inside a tissue and ranges from 2 to 10. Patients with a combined score between 2 and 4 mostly survive while patients with a score of 8 to 10 have a higher mortality rate.[2] Automating the cancer diagnosis and the grading process can reduce the time needed by pathologists and remove the inter- and intraobserver variations.[3]

Automated medical image classification is an important research area which utilizes different feature detection and representation techniques. These features can be classified into two main categories: hand-crafted and learned features. Hand-crafted features are based on the pathologists' approaches for cancer diagnosis and grading. Pathologists scan tissue slides and try to find symptoms of tumor progress including irregularly shaped nuclei and lack of differentiation. Therefore, most of automated histological analysis methods first segment the cell nuclei, then extract features from cell nuclei and use them for classification.[4–6] For example, Boucheron et al.[4] performed image segmentation on histopathology images of breast and used the extracted features for breast cancer detection. Farjam et al.[5] segmented the prostate glands and extracted structural features from them and used them in a tree-structured algorithm for automatic Gleason grading of prostate. Stotzka et al.[6] used features extracted from cell nuclei along with neural networks for automatic grading of the prostate. Some other techniques that used hand-crafted features are based on texture, color, and morphological features.[7–12] Jafari-Khouzani and Soltanian-Zadeh[7] extracted energy and entropy from multiwavelets coefficients and used them for the task of automatic Gleason grading of prostate. They used the k-nearest neighbor algorithm to classify images. Tabesh et al.[8] extracted color, texture, and morphometric features from microscopic images of prostate and combined them and used them for prostate cancer diagnosis and Gleason grading. In an earlier study,[9] we extracted features from multi-decomposition levels of shearlet filters and used the histogram of shearlet coefficients (HSCs) for the task of classification of benign and malignant breast slides using support vector machine

*Address all correspondence to: Hadi Rezaeilouyeh, E-mail: hadi. rezaeilouyeh@du.edu

(SVM) and achieved 75% classification accuracy. We also used HSC for prostate cancer detection of histological images[10] and magnetic resonance (MR) images[11] and achieved 100% and 97% classification accuracy, respectively. In a recent study,[12] we extracted multiple features from the histological images of prostate cancer and used the multiple kernel learning (MKL) algorithm for fusing the features. Then we used SVM equipped with MKL for the classification of prostate slides with different Gleason grades.

Most of the methods mentioned above merge different hand-crafted features to represent the texture of histopathology images. These methods usually include some preprocessing (e.g., segmentation) and the final classification result depends on the accuracy of the previous steps. In addition, these hand-crafted features are designed for a specific type of problem. It is time-consuming to design a new set of features and they are not easily applicable on different datasets. Furthermore, some of these hand-crafted features have inherent limitations that make them less efficient for complex tasks. A good example would be the wavelet transform, which has been widely used for different applications including cancer diagnosis and grading.[5–8] Wavelets do not have directional sensitivity, which makes them unsuitable for detecting directional features. That was the motive for using shearlets instead of wavelets in our previous studies[9–12] and in this paper as well. On the other hand, recent feature learning methods have gained a lot of attention due to the success of deep neural networks methods in computer vision applications. The advances in computational power and the availability of large training databases also played an important role in the development of deep neural networks.[13] Deep learning (DL) is a subset of machine learning models that represents high-level abstractions extracted directly from images using nonlinear transformations.[13,14] The main advantage of DL methods is their ability to form hierarchical representations of data by deriving higher level features from lower level features using nonlinear processing units.[13] Therefore, despite hand-crafted features, learned features do not need any preprocessing and can easily be transferred to different applications since they are data-driven.[15] These methods often outperform traditional approaches that use hand-crafted features.[16–19] Cruz-Roa et al.[15] proposed a three-layer convolutional neural network (CNN) method for invasive ductal carcinoma detection in histopathology images of breast cancer and compared their method with hand-crafted features. They reported 6% improvement in the classification accuracy when using their CNN. Liao et al.[16] proposed a stacked-independent subspace analysis DL framework for prostate T2 MR image segmentation. They reported 3% improvement in segmentation accuracy when using their DL framework. Couture et al.[17] proposed a sparse coding-based hierarchical feature learning method for breast cancer detection in histopathology images. They were able to increase the classification accuracy by 6% using their proposed feature learning method. Cireşan et al.[18] presented a DL method based on max pooling for mitosis detection in histology images of breast cancer. They won the International Conference on Pattern Recognition 2012 competition. Li et al.[19] used shearlet transform and deep neural networks for image quality assessment. They extracted features using the sum of subband shearlet coefficients and used stacked autoencoders as their main neural network building blocks. They used a softmax classifier to assess the quality of images in their dataset. In this paper, we propose a shearlet-based deep neural network method for breast cancer detection and Gleason grading of prostate. To the best of our knowledge, this is the first time that shearlet transform and DL are employed together for medical image classification.

Our main contribution in this paper is threefold. First, we propose using the phase of shearlet coefficients as a primary feature for general purpose microscopic medical image classification. This is the first study that utilizes the phase of shearlets for such applications. Shearlet transform[20] is a directional multi-scale representation system with affine properties which can detect anisotropic features at different orientations and scales. Most of the signal's information is carried by the phase[21] and the phase features are invariant to noise and image contrast.[22] However, since the phase information is nontrivial, it is difficult to design and hand-craft phase features that work as a general approach. This motivated us to further improve our proposed method as follows. Second, we add the magnitude of shearlet coefficients and the RGB images to the phase features. The magnitude of shearlet coefficients is a direct representative of the singularities in the image and the higher the magnitude, the higher the possibility of an edge occurring in that location.[23] The reason to include the RGB images is in the nature of this problem. Since these images are H&E stained, the color information is very important for breast cancer detection and Gleason grading. As cancer happens and the grade increases, the cell nuclei (stained blue) become larger and the cytoplasm area (stained pink) shrinks. Therefore, we need to consider color information as one of our primary features as well. Third, we propose a deep neural network as an evolution process to explore the aforementioned features (phase and magnitude of shearlet coefficients and RGB images) and use them for cancer detection and Gleason grading. Our CNN consists of multiple layers of convolutions followed by max pooling along with fully connected and dropout layers. In summary, our contributions are listed below:

- Utilize phase of shearlet coefficients as a primary feature for microscopic medical image classification.

- Empower magnitude of shearlet coefficients and RGB images to support the phase features.

- Employ deep neural networks to explore the shearlet-based image representations and RGB images and learn features for image classification.

The proposed framework for microscopic medical image classification is presented in Fig. 1. The remainder of this paper is organized as follows. The proposed framework consisting of shearlet transform and CNN is presented in Sec. 2. In Sec. 3, the experimental setup and results along with the analysis of the proposed framework are presented in detail. Finally in Sec. 4, discussions and conclusions are presented.

2 Methodology

Prostate Gleason grading and breast cancer detection are mainly based on texture features and characteristics of cancerous tissues as shown in Fig. 2. It is noticeable that as the Gleason grade increases [Fig. 2(b)], the texture becomes more detailed and the epithelial cell nuclei grow in a random manner and spread across the tissue. Therefore, we need an accurate and robust texture analysis technique. For this purpose, we propose to extract our primary features from microscopic images using magnitude and phase of shearlet coefficients and evolve these features using DL techniques to make them more discriminative. In
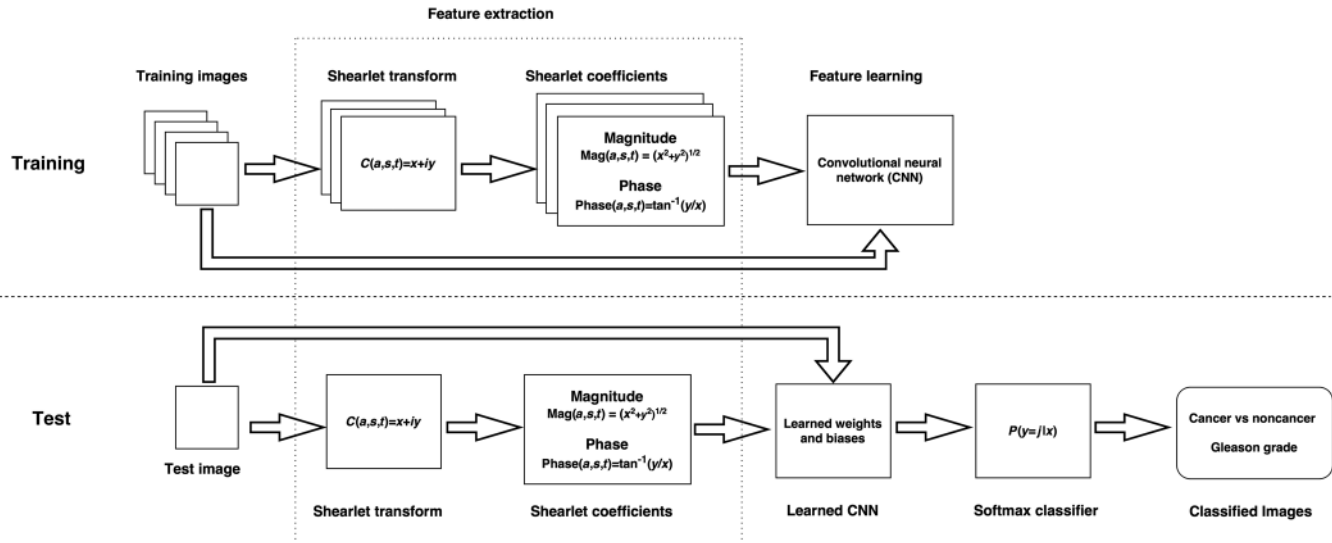
Fig. 1 Block diagram of our proposed framework consisting of the training and test phases.
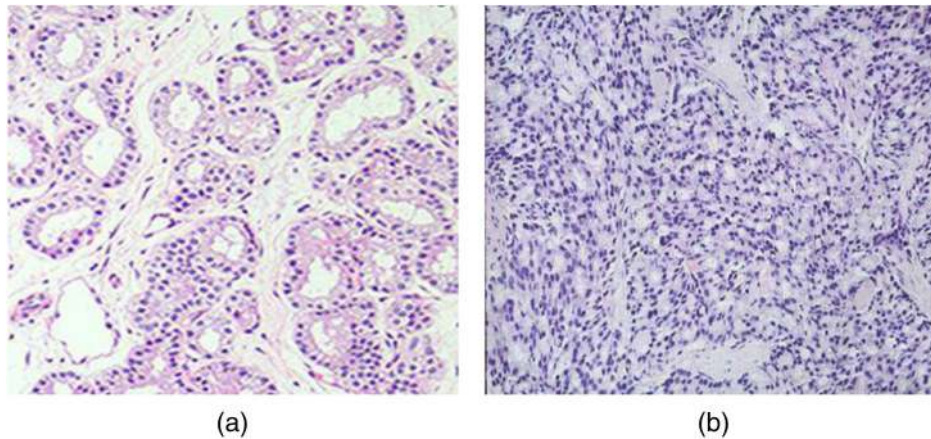

Fig. 2 Prostate tissue samples with different Gleason grade: (a) grade 2 and (b) grade 5.

the following subsections, we will describe both shearlet transform and DL in detail.

2.1 Shearlet Transform

Our proposed medical image analysis frameworks are based on features extracted from the shearlet transform.[20,24] Shearlet is a multiscale expansion of traditional wavelet transform that is efficiently developed to detect one- and two-dimensional (2-D) directional features in images. Despite its predecessors, e.g., curvelet[25] where the direction is defined using rotation, shearlet defines the direction using shearing matrices, which makes the discrete implementation of shearlets easier and also makes the shearlet rotation-invariant.

The continuous shearlet transform[20] is defined as the mapping for $f \epsilon R^2$

$$SH_\Psi f(a, s, t) \leq f, \Psi_{a,s,t} >, \quad a > 0, s \epsilon R, t \epsilon R^2, \quad (1)$$

where the shearlets are defined as the following:

$$\Psi_{a,s,t}(x) = |\det M_{a,s}|^{-\frac{1}{2}} \Psi(M_{a,s}^{-1} x - t), \quad (2)$$

where $M_{a,s} = \begin{pmatrix} a & \sqrt{a}s \\ 0 & \sqrt{a} \end{pmatrix} = B_s A_a$, where $A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$, and $B_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$. $M_{a,s}$ is the multiplication of an anisotropic dilation matrix (A_a) and a shearing matrix (B_s), which makes the shearlets well localized. By incorporating the translation, scale, and shear parameters, shearlet is able to detect directional singularities and geometrical features of multidimensional data.

A recent implementation of shearlets called "fast finite shearlet transform" (FFST)[24] uses fast Fourier transform for discrete implementation of shearlets in the frequency domain and consequently produces complex shearlet coefficients. Since we wanted to extract the phase of shearlet coefficients, we utilized this implementation of shearlets in this paper. The phase along with the magnitude of shearlet coefficients is then fed to a deep neural network for automatic Gleason grading and breast cancers diagnosis.

Shearlet has some interesting mathematical properties.[26,27] Shearlets are well localized; they are compactly supported in the frequency domain. They have parabolic scaling and each element of shearlets is supported on a pair of trapezoids. They have high directional sensitivity. Shearlets are spatially localized

and optimally sparse. To summarize, shearlets form a tight frame of well-localized waveforms at different scales and directions and are optimally sparse for representing edges in the images. These properties make the shearlet a well-suited tool for detecting singularities in different cancerous tissues in microscopic images.

2.2 Features Extracted from Shearlet Transform

The success of a classification framework highly depends on the choice of the feature representation method. In this paper, we are interested in microscopic image classification, especially in prostate Gleason grading and breast cancer diagnosis. Taking another look at Fig. 2 highlights the changes that a tissue cell

goes through when the Gleason grade increases. As the Gleason grade increases, epithelial cells randomly duplicate, disturbing the normal structure of glands.[4] The higher grade cells are described by irregular morphology in nuclei, larger nuclei, and less cytoplasm than lower grades as shown in Fig. 2. A similar process happens to breast tissue when cancer develops. To represent these textural and morphological properties of the cancerous tissues, we apply the shearlet transform on microscopic images and extract magnitude and phase of complex shearlet coefficients and use them as our primary features. To better illustrate the effectiveness of the shearlet transform for microscopic image classification, we show benign and malignant breast tissue images along with their corresponding magnitude and phase of shearlet coefficients from a single subband in Fig. 3. One can
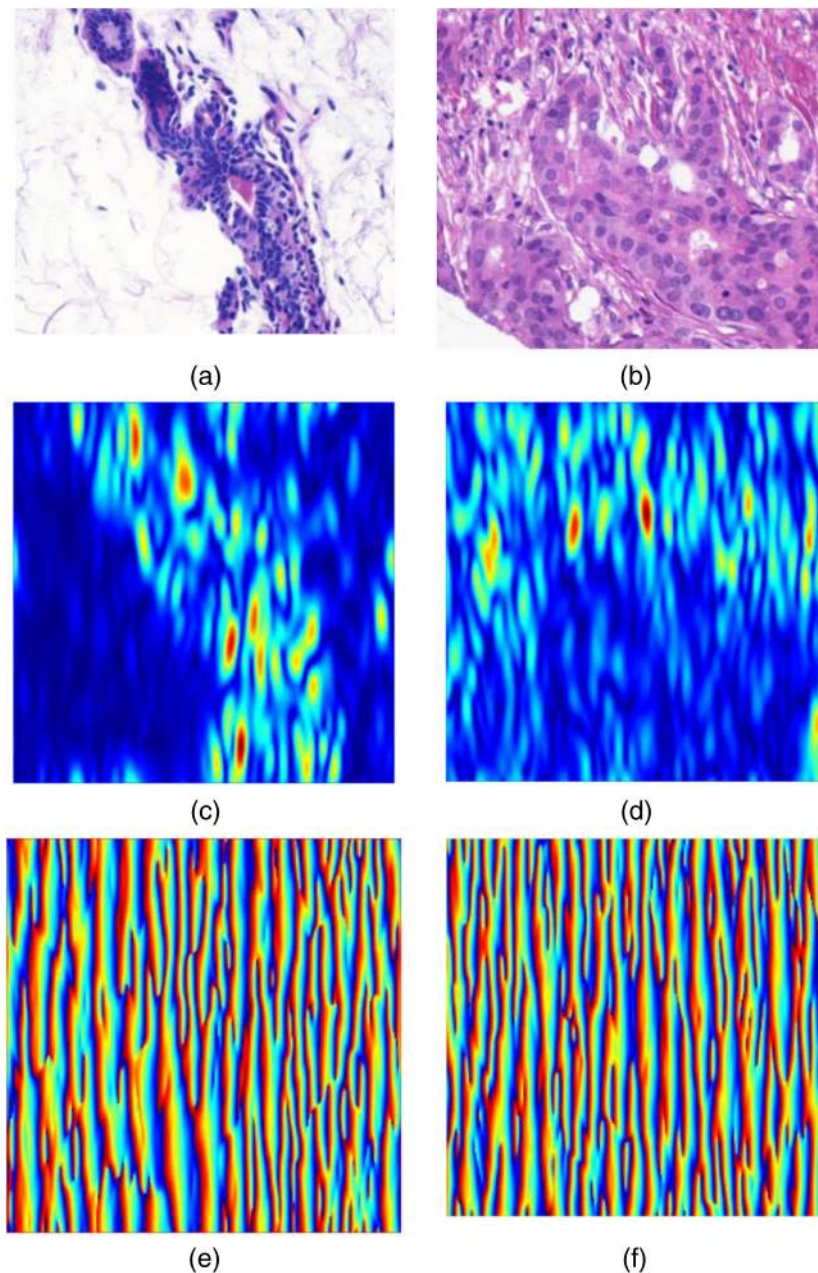
Fig. 3 Sample images of breast tissue and their corresponding magnitude and phase of shearlet coefficients from a single subband: (a) original benign, (b) original malignant, (c) magnitude of shearlets for benign, (d) magnitude of shearlets for malignant, (e) phase of shearlets for benign, and (f) phase of shearlets for malignant.
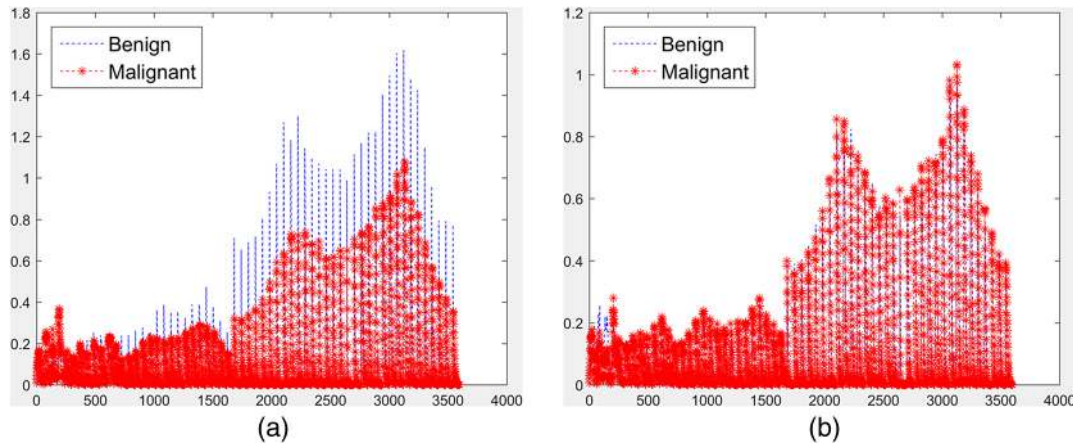
Fig. 4 HSCs for (a) correctly classified benign and malignant pair, and (b) incorrectly classified pair.

observe how the statistics of shearlet coefficients change as the tissue transforms from benign to malignant. This effect is more obvious in the magnitude of shearlet coefficients compared to the phase of shearlet coefficients. This is because the magnitude of shearlet coefficients is a direct representative of the singularities in the image and the higher the magnitude, the higher the possibility of an edge occurring in that location.[23] One way to represent statistical properties of shearlet coefficients is to extract histograms from the magnitude of shearlet coefficients. We previously used HSCs for breast cancer detection[9] and prostate cancer detection[11] and Gleason grading.[10,12] Figure 4 shows the histogram of magnitude of shearlet coefficients for two cases. Figure 4(a) shows the HSCs for a pair of benign and malignant images, where they were correctly classified. The shape and peaks of histograms are different for benign and malignant images. Figure 4(b) shows a failed case where the images were incorrectly classified using the HSC method. The shape and the peaks of the histograms are very similar. One possible reason is that the histogram does not include any information on the local structure of the images.

The importance of phase in image processing and computer vision has been investigated in previous studies.[28–22] It was verified that most of the signal's information is carried by the phase[21] and in some cases only the phase is enough to reconstruct a signal.[28] Also, the phase features are invariant to noise and image contrast.[22] However, since the phase information is nontrivial, it is difficult to design and hand-craft phase features that work as a general approach. For example, the histogram of phase features does not sufficiently represent the changes in the texture of an image since, despite magnitude, the phase does not directly relate to strong edges. This motivated us to add the phase information to the magnitude and learn features instead of hand-crafting them in this paper. We extract the magnitude and phase of shearlet coefficients as follows.

Assume we denote a complex shearlet coefficient by $c(a, s, t) = x + iy$, where x and y are the real and imaginary parts of a complex shearlet coefficient and a, s, and t are the scale, shear, and translation parameters of the shearlet transform, respectively. We use the following equations to extract the magnitude [$\text{mag}(a, s, t)$] and the phase [$\text{phase}(a, s, t)$] of the coefficients

$$\text{mag}(a, s, t) = |c(a, s, t)| = \sqrt{x^2 + y^2}$$

$$\text{phase}(a, s, t) = \measuredangle[c(a, s, t)] = \tan^{-1}\left(\frac{y}{x}\right). \qquad (3)$$

After we extracted the above features from each subband of the shearlets, then we feed them to our deep neural network as explained in Sec. 2.3.

2.3 Convolutional Neural Networks and Feature Learning

Our primary features are magnitude and phase of shearlet coefficients along with RGB data. As we previously explained, hand-crafting the features is not a suitable method for complex tasks (e.g., medical image analysis, where important clinical features are used to represent objects such as cell nuclei). Therefore, in the following, we propose our automatic feature learning method based on deep neural networks.

Traditional machine learning methods have limited abilities to analyze natural data that are represented in their raw form. This is due to the fact that shallow classifiers need appropriate feature extraction and representation techniques, which are sensitive to the discriminative attributes of the input while invariant to unimportant features (selectivity–invariance dilemma).[29] On the other hand, DL methods[29–32] consist of multiple processing layers to learn representations directly from raw data with multiple levels of abstraction.[29] For an image, the lower levels of abstraction might correspond to the edges in the image, while higher abstraction layers correspond to the objects in the image.

CNNs[30] are feed-forward networks consisting of consecutive pairs of convolutional and pooling layers along with fully connected layers. They are especially designed for inputs represented as 2-D data (e.g., images). The input data first go through pairs of convolution and pooling layers. Convolution layers apply 2-D convolution on their inputs using rectangular filters, which are applied in different positions of the input. The convolution layer sums the responses from previous layer, adds a bias term, and drives the result through a nonlinear activation function. This process is repeated with different weights to create multiple feature maps. The output of the convolutional layer then usually passes through a pooling layer, which is a downsampling technique and results in translation-invariant features. After a few pairs of convolution and pooling layers, one or more fully connected layers combine the outputs into a feature vector. The final layer is a fully connected layer with one neuron per class (two for breast cancer diagnosis and four for Gleason grading), which are activated by a softmax classifier. Throughout the whole process, the weights are optimized by minimizing the misclassification error using stochastic gradient descent method.
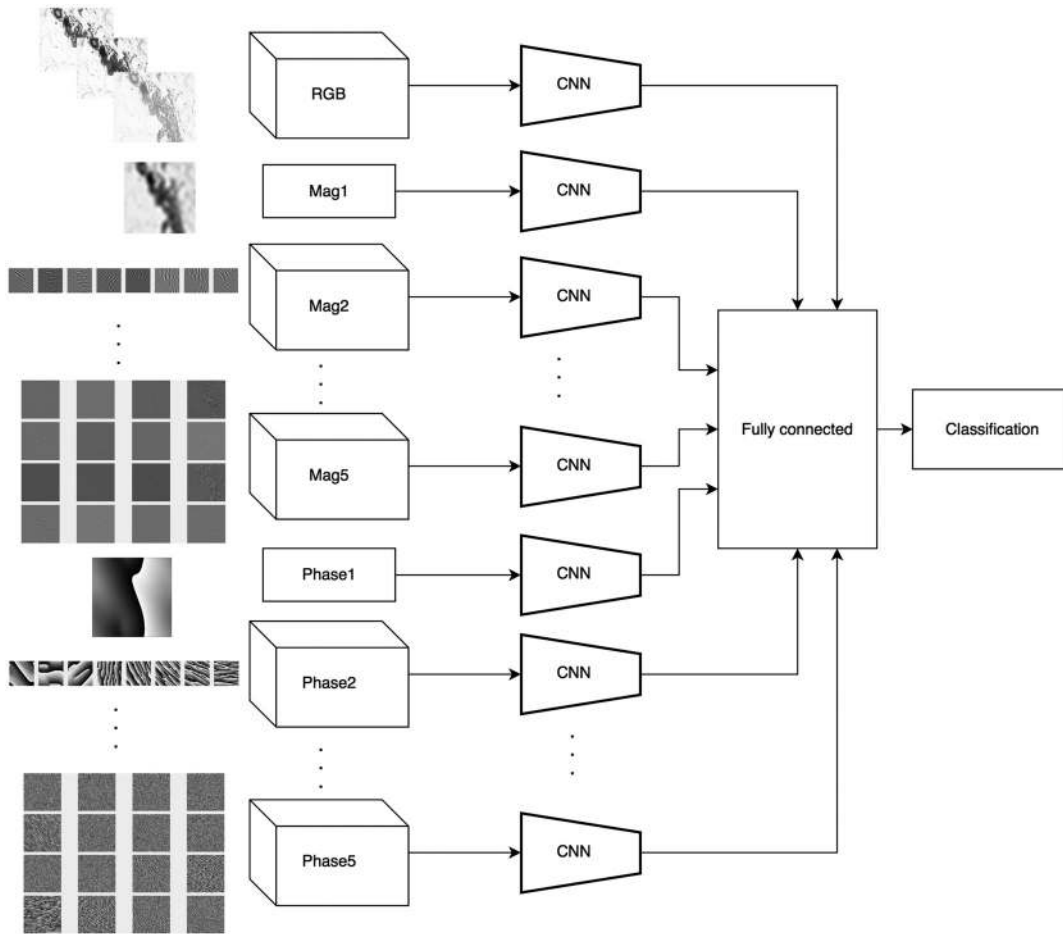
Fig. 5 Block diagram of our deep neural network. The inputs are RGB images, magnitude of shearlet coefficients from decomposition levels 1 to 5 (Mag1 to Mag5), and phase of shearlet coefficients from decomposition levels 1 to 5 (Phase1 to Phase5). Then they go through separate CNNs and the results are concatenated using a fully connected layer, which sends the final evolved features to softmax for classification.

The building blocks of our proposed deep neural network are presented in Fig. 5. Figure 6 shows the architecture of our CNN. The description of each layer in our CNN is presented in the following:

1. Convolutional layer (conv): This layer applies a 2-D convolution on the input feature maps using 64 Gaussian filters of size 5×5 initialized with a standard deviation of 0.0001 and bias of zero. It steps 2 pixels between each filter application. The output then goes through a nonlinear rectified linear unit (ReLU) function, which is defined as $f(z) = \max(z, 0)$. This nonlinear activation function is important since it lets the network learn abstracts using a small number of nodes. Otherwise,
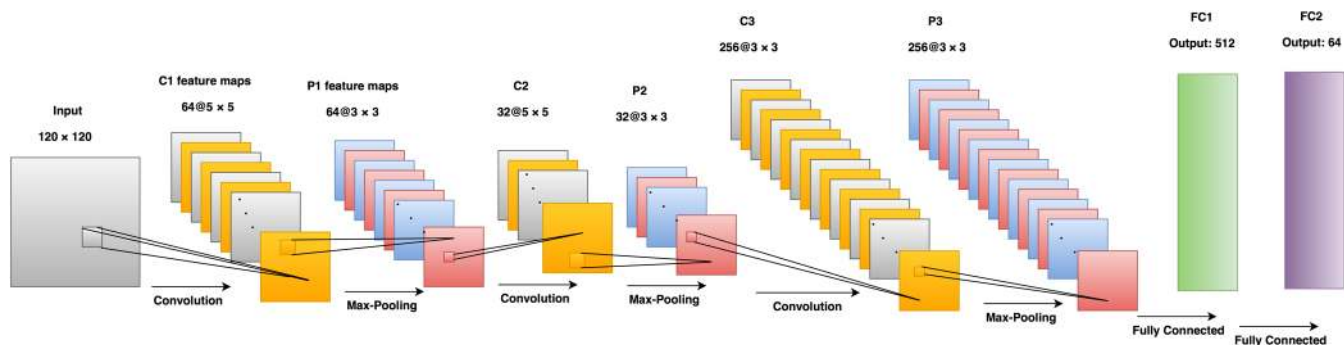
Fig. 6 Architecture of our CNN. The input is a 120×120 patch and can be either RGB or magnitude or phase of shearlet coefficients. Then three layers of convolution and pooling are applied on the input back to back to extract abstracts from the input. Finally, a fully connected layer combines the outputs of convolution filters and sends out a single feature vector with the size of 64.

if a linear function was used, the entire network would be equivalent to a single-layer neural network.

2. Max-pooling layer: The purpose of the pooling layer is to combine similar features into one; therefore, it is a feature dimension reduction technique. It calculates the maximum of a local patch of units inside a 3×3 region of input feature map and steps 2 pixels between pooling regions. This makes the learned features invariant to shifts and distortions.

3. Local response normalization (LRN): Performs normalization on local input regions by dividing each input by $[1 + \frac{\alpha}{n}(\sum_i x_i^2)]^\beta$, where x_i is the i'th input, $n = 3$ is the size of local region, $\alpha = 5 \times 10^{-5}$, and $\beta = 0.75$.

4. Fully connected layer: Also known as inner product this is one of the top layers of a CNN architecture. It connects all the neurons in the previous layer to all the neurons it has. It does not have any spatial information. It takes the output of pooling layers as input and combines them into a feature vector, similar to a multilayer perceptron network.

5. Dropout: Dropout regularization[33] randomly disables portion of neurons in the training set. This prevents the overfitting of learning results to the structure of the network. In this paper, we chose the dropout threshold to be 0.7.

6. Classification layer: The final layer of CNN is a fully connected layer with one neuron per class (two for breast cancer diagnosis and four for Gleason grading), which is activated by the softmax classifier. Given the training labels, it gives the accuracy of the classification.

3 Experiments and Results

In this section, first we describe our data preprocessing in detail. Then we explain our feature extraction using shearlets transform and CNN structure and parameters. Finally, we present our results and compare them with the state-of-the-art methods based on hand-crafted features.

3.1 Datasets and Data Preparation

We used two microscopic medical imaging datasets for our experiments. The first set was the University of California, Santa Barbara Biosegmentation Benchmark dataset.[34] This dataset contained 58 H&E-stained histopathology images of breast tissue. Out of 58 total images, there were 26 malignant images and 32 benign images. The second dataset was the prostate Gleason grading dataset used by Jafari-Khouzani and Soltanian-Zadeh.[7] This dataset contained 100 H&E images of prostate tissue samples. The images were of grades 2 to 5 and the magnification was 100 with different sizes. All of the images were captured in equal conditions of light. This dataset contained 21, 20, 32, and 27 images of grades 2, 3, 4, and 5, respectively. Each image had a single grade. The images were graded by expert pathologists who provided the ground truth data.

Since our CNN experiments needed a large amount of data, we augmented both datasets. For this purpose, we performed mirroring, patches, rotation, and scaling of the images. For mirroring, we used three mirroring scenarios (horizontal, vertical, and horizontal and vertical). For rotation, we rotated each image 10 times with a rotation randomly chosen between 10 deg and 90 deg. For scaling, we resized each image by a factor of 2. For extracting patches out of images, we extracted them from top left, top right, bottom left, bottom right, and center of the image, each half the size of the original image. We also combined the above operations to further augment the datasets. Overall, we were able to augment each original image to 104 images. Therefore, we had 6032 augmented breast tissue images and 10,400 augmented prostate tissue images. Throughout this whole process, since we had different size images, we resized the images to 128×128 pixels for normalization purposes. Figure 7 shows all 104 augmented images of a sample breast tissue image.

3.2 Experimental Setup

We had two types of primary features. One was the RGB images, which were extracted as explained in Sec. 3.1. The other primary features were shearlet features, which were extracted as explained next.

3.2.1 Shearlet feature extraction

To apply shearlet transform on images, we utilized the FFST MATLAB® toolbox provided by Häuser and Steidl.[24] We chose five scales (decomposition levels) for shearlet. The first decomposition level was a low-pass filtered version of input. We chose eight directions for the second and third levels and 16 directions for the fourth and fifth levels which led to 8, 8, 16, and 16 subbands, respectively. Therefore, overall we had $1 + 8 + 8 + 16 + 16 = 49$ subbands of shearlets. All these subbands were of the same size as the input image (150×150). We followed the procedure explained in Sec. 3.1 to extract the magnitude [$\mathrm{mag}(a, s, t)$] and phase [$\mathrm{phase}(a, s, t)$] of shearlet coefficients from each subband and fed them to our CNN framework.

3.2.2 Convolutional neural networks framework and feature evolution

As we explained in Sec. 2.3, our CNN consisted of three layers of convolution and max-pooling. For convolutional layers, we initialized 64 Gaussian filters of size 5×5 with a standard deviation of 0.0001 and bias of zero. The step between each filter application was 2 pixels. We used an ReLU function as the activation function. For max-pooling layer, we applied it on local patch of units inside a 3×3 region of input feature map with a 2 pixels step between pooling regions. We used an LRN layer to normalize local input regions. We used fully connected layers for concatenating the outputs of CNNs.

We used the stochastic gradient descent algorithm with the momentum of 0.9 and the weight decay of 0.05 in all experiments. We used mini-batches of 32 samples due to the large size of the network and the memory limitations. All models were initialized with the learning rate of 0.001. These hyperparameters were empirically found based on the performance of validation set over onefold of Gleason grading. The same hyperparameters were used for the breast cancer experiment.

We also used dropout layers to prevent the overfitting of the results to the structure of the CNN. The dropout threshold value

Fig. 7 Augmented images of a sample breast tissue image from our dataset.

of 0.7 was found to be the best based on the classification accuracy on validation set.

Our primary input features were RGB, magnitude, and phase of shearlet coefficients. Figure 5 shows the overall structure of our deep neural network. Mag1 to Mag5 were the magnitude of shearlet coefficients from decomposition levels 1 to 5, respectively. Phases 1 to 5 were the phase of shearlet coefficients from decomposition level 1 to 5, respectively. We fed each one of our inputs (RGB, Mag1 to Mag5, and phases 1 to 5) to a separate CNN. The reason for separating RGB from shearlet data was because they were of a different nature and, therefore, needed separate processing. We separated the magnitude and phase of shearlets for the same reason. We also processed shearlet coefficients from different decomposition levels independently because different decomposition levels represent features from different scales.

Figure 8 visualizes the shearlet feature evolution as they go through each convolutional layer. Figures 8(a) and 8(b) show the first convolution layer output features for the first and third decomposition level shearlet coefficients, respectively. The third decomposition level shearlet coefficients represent more details in the images with more directional sensitivity. Figures 8(c) and 8(d) show the same shearlet coefficients out of the second convolution. After the second convolution, the features become more distinguishable.

3.3 Results

We evaluated our proposed microscopic image classification framework for two tasks: breast cancer diagnosis and prostate Gleason grading. Although both tasks contain similar input data (H&E images), they are different in nature. One is to distinguish cancerous from noncancer cells, while the other (i.e., Gleason grading) is to evaluate how advanced the cancer is. Also, they belong to different human tissues, therefore, the physiological and textural information are different. We evaluated our method against these different tasks to show the generality and applicability of our method.

For each classification task, RGB images and extracted shearlet features from input images were fed to our CNN framework with the parameters explained in Sec. 3.2.2. For cross validation, we used a fivefold cross-validation technique. We divided our original datasets (nonaugmented) into five sets and used four sets for training and one for testing. We repeated this five times and reported the average classification accuracy. We used the augmentation process during the training. The final network is evaluated on the original images only. Therefore, all images pertaining to a given case are either in the training or test set (not in both). We had three different scenarios for CNN experiments. In the first scenario, we used only RGB data as input. In the second scenario, we combined RGB and magnitude of shearlets and used them as input. Lastly, we combined RGB, magnitude, and phase of shearlets and used them as input to CNN. This helped us understand the contribution of each feature set separately and when combined together. We were able to significantly increase the classification accuracy (by 13% for breast cancer diagnosis and 8% for Gleason grading) by combining RGB and magnitude of shearlets. We further improved the results by including phase information as well. To evaluate the performance of our deep neural network, we compared the results with the state-of-the-art methods based on hand-crafted features using SVM. For SVM, we tried different kernels (linear, polynomial, and RBF) with different parameters (polynomial order of 1, 2, and 3 for polynomial kernel and sigma values
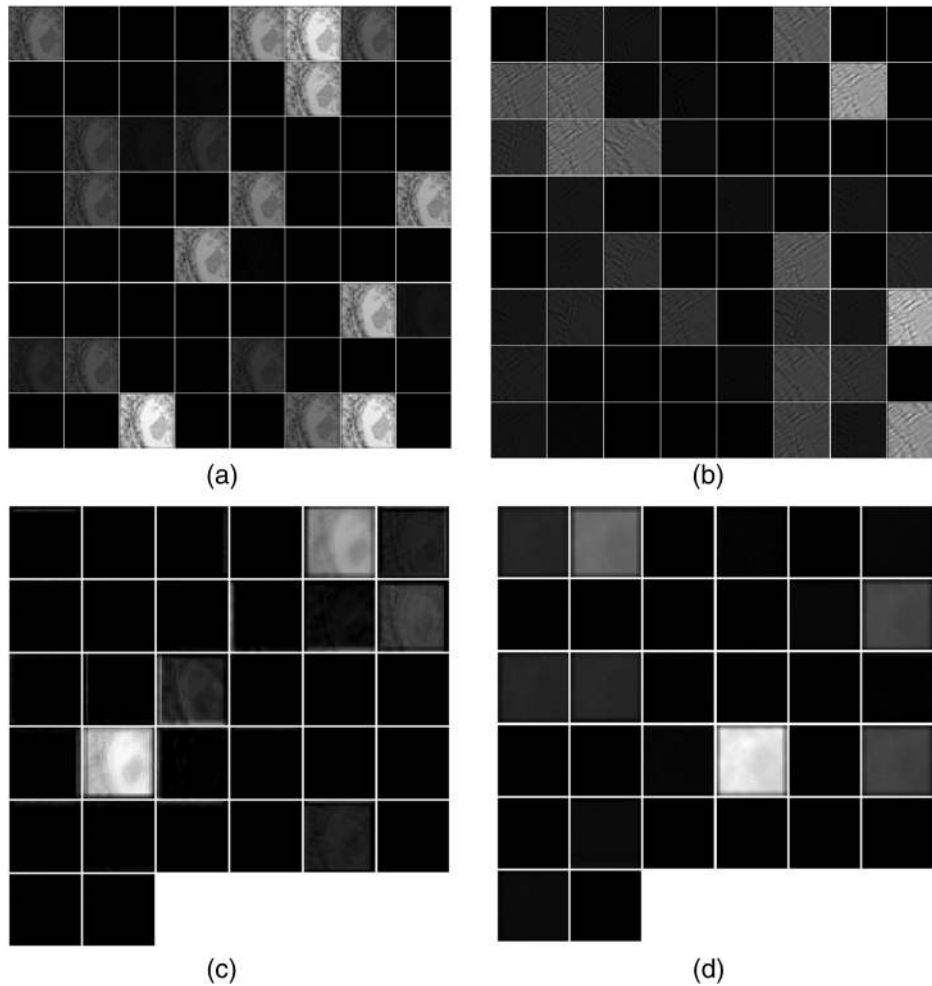
Fig. 8 Feature evolution: (a) first convolutional layer output features for magnitude of shearlet coefficients from first decomposition level, (b) first convolutional layer output features for magnitude of shearlet coefficients from third decomposition level, (c) second convolutional layer output features for magnitude of shearlet coefficients from first decomposition level, and (d) third (last) convolutional layer output features for magnitude of shearlet coefficients from third decomposition level.

between 1 and 10,000 for Gaussian radial basis function kernel) and chose the best kernel and parameters for each experiment. Our experiments showed that CNN outperforms the hand-crafted feature extraction methods.

Table 1 shows the classification results for breast cancer diagnosis using our deep neural network method and state-of-the-art methods. In addition to the classification accuracy, we have also reported the sensitivity, specificity, F-1 score, and area under the curve (AUC) as performance metrics. Table 1 shows the

average values of the each metric along with the standard deviation values over fivefolds. It is obvious from the table that by including the magnitude and phase of shearlet coefficients we achieved higher performance metrics. Table 1 also shows the breast cancer classification results using hand-crafted features. These results show the superiority of our proposed method over hand-crafted feature extraction methods for breast cancer detection.

Table 2 shows the classification results for automatic Gleason grading using our deep neural network method and

Table 1 Classification results for breast cancer detection (mean ± std).

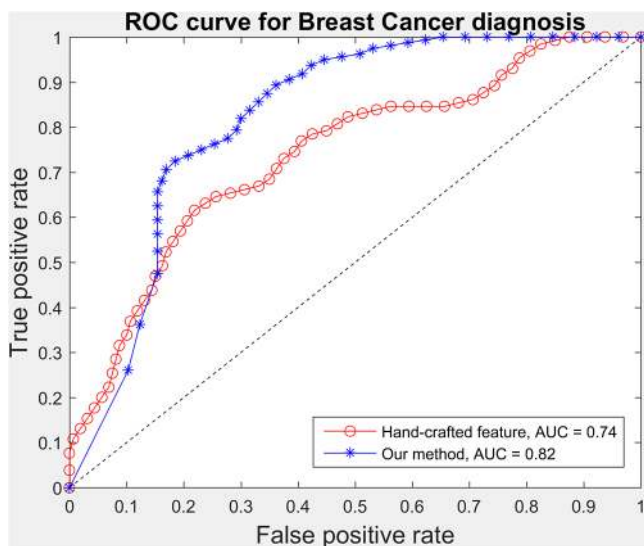| Method	Sensitivity	Specificity	F-1 Score	AUC	Accuracy
RGB	0.91 ± 0.08	0.59 ± 0.09	0.76 ± 0.05	0.68 ± 0.02	0.71 ± 0.02
RGB + magnitude of shearlets	1	0.62 ± 0.10	0.84 ± 0.03	0.78 ± 0.01	0.84 ± 0.01
RGB + magnitude + phase of shearlets	**1**	**0.72 ± 0.10**	**0.89 ± 0.03**	**0.82 ± 0.01**	**0.86 ± 0.03**
Boucheron et al.[4]	—	—	—	—	0.74
Rezaeilouyeh et al.[9]	0.93 ± 0.09	0.60 ± 0.10	0.79 ± 0.06	0.74 ± 0.02	0.74 ± 0.09

Note: Bold values indicate the best results.

Table 2 Classification results for Gleason grading (mean ± std).

Method	Sensitivity	Specificity	F-1 Score	AUC	Accuracy
RGB	0.80 ± 0.02	0.91 ± 0.01	0.71 ± 0.01	0.72 ± 0.02	0.76 ± 0.06
RGB + magnitude of shearlets	0.84 ± 0.01	0.91 ± 0.02	0.81 ± 0.03	0.79 ± 0.02	0.84 ± 0.04
RGB + magnitude + phase of shearlets	**0.89 ± 0.01**	**0.94 ± 0.01**	**0.85 ± 0.02**	**0.84 ± 0.01**	**0.88 ± 0.05**
Jafari-Khouzani and Soltanian-Zadeh[7]	0.82 ± 0.01	0.91 ± 0.02	0.73 ± 0.02	0.78 ± 0.02	0.83 ± 0.09
Rezaeilouyeh et al.[10]	0.78 ± 0.03	0.91 ± 0.01	0.69 ± 0.03	0.74 ± 0.01	0.78 ± 0.11
Wavelet packet[35]	0.82 ± 0.02	0.92 ± 0.01	0.73 ± 0.01	0.74 ± 0.02	0.78 ± 0.07
Co-occurrence matrix[36]	0.81 ± 0.01	0.92 ± 0.01	0.72 ± 0.02	0.73 ± 0.02	0.77 ± 0.09

Note: Bold values indicate the best results.

state-of-the-art hand-crafted feature extraction methods. Similar to breast cancer diagnosis, by including the magnitude and phase of shearlet coefficients we improved the classification accuracy. Table 2 also shows the Gleason grading classification results using hand-crafted features. For Jafari-Khouzani and Soltanian-Zadeh,[7] we extracted the energy and entropy of multi-wavelets and used them as features in the SVM classifier. For the wavelet packet[35] and co-occurrence matrix[36] features, we used MATLAB®.[37] We used the MATLAB's® image processing and wavelet toolbox for this purpose. These results show the advantage of our proposed method over state-of-the-art hand-crafted feature extraction methods for Gleason grading.

The receiver operating characteristic (ROC) curve for breast cancer diagnosis is shown in Fig. 9. In this figure, we compare hand-crafted feature extraction method[9] with the best results from our deep CNN method. An ROC curve depicts the true positive rate against the false positive rate for different thresholds. It can be observed from Fig. 9 and based on their AUC values that our CNN method outperforms the best hand-crafted feature extraction method.[9]

We also report the confusion matrix (%) for the automatic Gleason grading experiments in Tables 3 and 4. Table 3 shows the confusion matrix for Gleason grading using the best hand-crafted feature extraction method,[7] while Table 4 shows the confusion matrix using our best CNN-based method. A confusion matrix is a table that is used to visualize the performance of a classifier using true and predicted labels. Since we have four classes in Gleason grading (grades 1 to 4), our confusion matrix is 4 × 4. It is noticeable that using our proposed method the misclassified cases only belong to Gleason grade 5. This is in accordance with the pathologists diagnosis since distinguishing grade 5 from grade 4 is the most difficult task in Gleason grading.[8] Our CNN method is 15% better than hand-crafted

Table 3 Confusion matrix (%) for Gleason grading experiment using the best hand-crafted feature extraction method[7].

		Grade 2	Grade 3	Grade 4	Grade 5
	Grade 2	100	0	0	0
True label	Grade 3	0	100	0	0
	Grade 4	0	3	97	0
	Grade 5	0	33	26	41
			Predicted label		

Table 4 Confusion matrix (%) for Gleason grading experiment using our deep neural network.

		Grade 2	Grade 3	Grade 4	Grade 5
	Grade 2	100	0	0	0
True label	Grade 3	0	100	0	0
	Grade 4	0	0	100	0
	Grade 5	7	0	37	56
			Predicted label		

Fig. 9 ROC curves for breast cancer diagnosis experiment using the best hand-crafted feature extraction method[9] and our best deep neural network results.

features in distinguishing between grades 4 and 5 (41% versus 56%). In addition, using hand-crafted feature extraction methods,[7] there are some misclassifications between grades 4 and 3 in addition to grades 4 and 5, which proves the advantage of our method.

4 Discussions and Conclusions

Early diagnosis of cancer and grading its severity are very important tasks that can save a patients' life. Automating this process can help pathologists to have a faster and more reliable diagnosis. Most of the automatic cancer diagnosis and grading techniques use hand-crafted features that need to be fine-tuned for different tasks.

In this paper, we proposed a framework for automatic breast cancer detection and prostate Gleason grading. First, we extracted the magnitude and phase of complex shearlet coefficients from the histological images. Shearlet transform is a multiscale directional system that has proven itself suitable for texture analysis of microscopic images in our previous studies. Then we combined the shearlet features with imagery data and used them to train CNNs. This feature learning process further enhanced the features and made them more discriminative. Then we used softmax classifier to distinguish different microscopic images. We were able to achieve high-classification accuracy on both breast cancer and Gleason grading datasets using our proposed method. We also compared our method against state-of-the-art methods that use hand-crafted features. We were able to outperform those methods in both cases.

One of the main advantages of our method is that it does not make any assumptions beforehand about the visual features of cancerous tissues. We consider shearlet transform as a general mathematical tool and extract features without any hand-crafting. Our deep neural network takes care of the feature learning task. Future work includes exploring the possibility of using deeper architectures for CNN and also expanding the applications of our method to different medical image analysis tasks.

References

1. American Cancer Society, Cancer facts and figures, Technical Report (2016).
2. D. F. Gleason, "Histologic grading of prostate cancer: a perspective," *Human Pathol.* **23**(3), 273–279 (1992).
3. C. Demir and B. Yener, "Automated cancer diagnosis based on histopathological images: a systematic survey," Rensselaer Polytechnic Institute, Technical Report (2005).
4. L. E. Boucheron, B. S. Manjunath, and N. R. Harvey, "Use of imperfectly segmented nuclei in the classification of histopathology images of breast cancer," in *2010 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 666–669, IEEE (2010).
5. R. Farjam et al., "Tree-structured grading of pathological images of prostate," *Proc. SPIE* **5747**, 840–851 (2005).
6. R. Stotzka et al., "A hybrid neural and statistical classifier system for histopathologic grading of prostatic lesions," *Anal. Quant. Cytol. Histol.* **17**(3), 204–218 (1995).
7. K. Jafari-Khouzani and H. Soltanian-Zadeh, "Multiwavelet grading of pathological images of prostate," *IEEE Trans. Biomed. Eng.* **50**(6), 697–704 (2003).
8. A. Tabesh et al., "Multifeature prostate cancer diagnosis and Gleason grading of histological images," *IEEE Trans. Med. Imaging* **26**(10), 1366–1378 (2007).
9. H. Rezaeilouyeh et al., "A microscopic image classification method using shearlet transform," in *2013 IEEE Int. Conf. on Healthcare Informatics (ICHI)*, pp. 382–386 (2013).
10. H. Rezaeilouyeh et al., "Prostate cancer detection and Gleason grading of histological images using shearlet transform," in *2013 Asilomar Conf. on Signals, Systems and Computers*, pp. 268–272 (2013).
11. H. Rezaeilouyeh et al., "Diagnosis of prostatic carcinoma on multiparametric magnetic resonance imaging using shearlet transform," in *2014 36th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 6442–6445 (2014).
12. H. Rezaeilouyeh and M. H. Mahoor, "Automatic Gleason grading of prostate cancer using shearlet transform and multiple kernel learning," *J. Imaging* **2**(3), 25 (2016).
13. Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.* **2**(1), 1–127 (2009).
14. Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives." *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1798–1828 (2013).
15. A. Cruz-Roa et al., "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," *Proc. SPIE* **9041**, 904103 (2014).
16. S. Liao et al., "Representation learning: a unified deep learning framework for automatic prostate MR segmentation," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, pp. 254–261, Springer, Berlin, Heidelberg (2013).
17. H. D. Couture et al., "Hierarchical task-driven feature learning for tumor histology," in *2015 IEEE 12th Int. Symp. on Biomedical Imaging (ISBI)*, pp. 999–1003 (2015).
18. D. C. Cireşan et al., "Mitosis detection in breast cancer histology images with deep neural networks." in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, pp. 411–418, Springer, Berlin, Heidelberg (2013).
19. Y. Li et al., "No-reference image quality assessment with shearlet transform and deep neural networks," *Neurocomputing* **154**, 94–109 (2015).
20. G. Easley, D. Labate, and W.-Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Appl. Comput. Harmon. Anal.* **25**(1), 25–46 (2008).
21. N. Skarbnik, Y. Y. Zeevi, and C. Sagiv, "The importance of phase in image processing," Technical Report, pp. 1–30 (2010).
22. P. Kovesi, "Phase congruency detects corners and edges," in *The Australian Pattern Recognition Society Conf.: DICTA* (2003).
23. G. R. Easley, D. Labate, and F. Colonna, "Shearlet-based total variation diffusion for denoising," *IEEE Trans. Image Process.* **18**(2), 260–268 (2009).
24. S. Häuser and G. Steidl, "Fast finite shearlet transform," arXiv preprint arXiv:1202.1773 (2012).
25. E. Candes et al., "Fast discrete curvelet transforms," *Multiscale Model. Simul.* **5**(3), 861–899 (2006).
26. P. Grohs et al., "Parabolic molecules: curvelets, shearlets, and beyond," in *Approximation Theory XIV: San Antonio 2013*, pp. 141–172, Springer International Publishing, Switzerland (2014).
27. G. R. Easley, D. Labate, and W. Q. Lim, "Optimally sparse image representations using shearlets," in *2006 Fortieth Asilomar Conf. on Signals, Systems and Computers*, pp. 974–978, IEEE, (2006).
28. A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE* **69**(5), 529–541 (1981).
29. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436–444 (2015).
30. Y. LeCun et al., "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**(11), 2278–2324 (1998).
31. I. Arel, D. C. Rose, and T. P. Karnowski, "Deep machine learning-a new frontier in artificial intelligence research [research frontier]," *IEEE Comput. Intell. Mag.* **5**(4), 13–18 (2010).
32. J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks* **61**, 85–117 (2015).

33. G. E. Hinton et al., "Improving neural networks by preventing co-adaptation of feature detectors," arXiv preprint arXiv:1207.0580 (2012).
34. E. D. Gelasca et al., "Evaluation and benchmark for biological image segmentation," in *IEEE Int. Conf. on Image Processing*, pp. 1816–1819, IEEE Signal Processing Society, San Diego, California (2008).
35. A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(11), 1186–1191 (1993).
36. R. M. Haralick and K. Shanmugam, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **6**, 610–621 (1973).
37. MATLAB Release 2015b, MathWorks, Inc., Natick, Massachusetts.

Hadi Rezaeilouyeh is a PhD student in electrical and computer engineering at the University of Denver. His research includes medical image analysis, machine learning, and computer vision.

Ali Mollahosseini is a PhD student in electrical and computer engineering at the University of Denver. His research includes machine learning, robotics, and computer vision.

Mohammad H. Mahoor is an associate professor of electrical and computer engineering at the University of Denver and the director of the Computer Vision Laboratory. He received his PhD from the University of Miami, Florida in 2007. His research includes visual pattern recognition, social robot design, and bioengineering.