Glandular Structure-Guided Classification of Microscopic Colorectal Images using Deep Learning

Ruqayya Awan[a,b,*], Somaya Al-Maadeed[a], Rafif Al-Saady[c], Ahmed Bouridane[d]

[a] *Department of Computer Science and Engineering, Qatar University, Doha, Qatar*
[b] *Department of Computer Science, The University of Warwick, Coventry, UK*
[c] *Al-Ahli Hospital, Doha, Qatar*
[d] *Computer and Information Sciences, Northumbria University, UK*

Abstract

In this work, we propose to automate the pre-cancerous tissue abnormality analysis by performing the classification of image patches using a novel two-stage convolutional neural network (CNN) based framework. Rather than training a model with features that may correlate among various classes, we propose to train a model using the features which vary across the different classes. Our framework processes the input image to locate the region of interest (glandular structures) and then feeds the processed image to a classification model for abnormality prediction. Our experiments show that our proposed approach improves the classification performance by up to 7% using CNNs and more than 10% while using texture descriptors. When testing with gland segmented images, our experiments reveal that the performance of our classification approach is dependent on the gland segmentation approach which is a key task in gland structure-guided classification.

Keywords: Colorectal cancer, glandular structures, gland segmentation, gland-guided classification, deep learning.

*Corresponding author

Email addresses: r.awan.1@warwick.ac.uk (Ruqayya Awan), S_alali@qu.edu.qa (Somaya Al-Maadeed), rafif.alsaady@gmail.com (Rafif Al-Saady), ahmed.bouridane@northumbria.ac.uk (Ahmed Bouridane)

This work is supported by a grant from the Qatar National Research Fund through National Priority Research Program (NPRP) No. 6-249-1-053.

1. Introduction

Adenocarcinoma is a prevalent type of colorectal cancer which appears to arise from the glands in the epithelial tissue and affects various body organs, including colon, prostate, lungs and breast. It is believed that adenocarcinoma occurs due to various genetic changes affecting the growth mechanism of cells [1] which causes the glandular structure to lose their normal morphological properties. In this study, our focus is on the colorectal adenocarcinoma (CRA) which accounts for more than 90% of colorectal cancer [2] and arises from the lining of the colon wall. In the case of adenocarcinoma, the glandular structures are currently considered as one of the important biomarkers for tumour grade determination. Normal glands appear to be in circular or ellipsoidal shape depending on the cutting plane. While the shape of malignant glands would deviate from the normal shape and the extent of the deviation would define the grade of the tumour to be either of a low grade or a high grade tumour in the two-tier grading system.

The advent of whole slide scanners has promoted the digital pathology in the clinical and research community. In the literature, most of the work on the development of an automatic system is focused on differentiating between benign/normal and malignant images while some for identifying different grades of cancer. In computational pathology, the identification of non-cancerous abnormal tissue is a challenging task due to their similarity with the normal and cancerous tissue. A very few works can be found on the identification of other abnormalities which are non-cancerous [3, 4] but in clinical practice, are substantial to locate as they can lead to carcinoma. In this study, we present an autonomous framework where stress is given to the abnormal tissue due to their ability to developing cancer and the need to learn distinctive features to increase the inter-class variability.

In this work, we consider pixel-level features to discriminate between normal, abnormal and cancerous images. These pixel-level features are learned from the glandular structures while ignoring the non-glandular tissue region. The

contributions of this paper are three-fold: (i) we propose a novel framework that harnesses gland-guided criteria of classification, by extracting features from the most discriminatory region of interest (ROI), (ii) we collected a colorectal image dataset, of medically significant tissue classes using our halogen-microscopic visual field acquisition system, and iii) extensive evaluation of our proposed method using a number of hand-crafted and CNN learned features. The scope of this study does not fall within the range of a mere application of CNN for the classification of histology images. Its purpose is to build a novel CNN based framework inspired by the histological need of the data under study. The benefits of our approach are experimentally verified using different types of features and this approach can be applied to other organ's cancerous and pre-cancerous tissues (breast and lung etc) possessing glandular structures.

The next section surveys the related work on classification and segmentation of histology images. In section 3, the clinical aspect of our dataset is explained. Section 4 presents our proposed framework while in section 5, we give details of our dataset and evaluation measures and present results along with the discussion. Finally, we conclude our study in section 6.

2. Related Work

In literature, studies on the classification of adenoma and adenocarcinoma can be divided into three categories: one which employs morphological features of the glands and its components while the second category uses the texture features and the third category makes use of deep learning approaches for stratification. One of the earliest work on morphological features [5] formulates an index based on the extent of variation in size and illumination of glands and its roundness to differentiate between benign and malignant prostate images. Naik et al. [6] performed three two-class classification of prostate tissue images: benign vs grade-3 cancer, benign vs grade-4 cancer and grade-3 vs grade-4 cancer. In this study, a set of various morphological features extracted from the boundary of the detected lumen and nuclei are used for classification. In [7], Nguyen

et al. presented an approach to classify images into benign, grade 3 carcinoma and grade 4 carcinoma of prostate. Their classification approach considers the overall structure of the gland by extracting the features from its various components: lumen, nuclei surrounding the gland, blue mucin and overall gland itself. Peng et al. [8] extracted morphological features of glands including the size, circularity and density of gland distribution to distinguish between malignant and non-malignant prostate images. Gultekin et al. presented a two-tier decomposition of tissue for the classification of colon tissue images. They formulated a new metric to characterize the glandular morphology and is referred to as *dominant blob scale* [9]. For classification, they extract graph-based features from the glandular objects. In another study [10], authors proposed a novel metric, Best Alignment Metric (BAM) to compute the distance between the shape of glands with a normal gland shape. This study demonstrated a strong correlation between the grade of colorectal cancer and BAM based features.

In [11, 3, 4, 12], several texture features have been tested for a better representation of colorectal histology images. In [13], Chaddad et al. used a set of three texture features extracted from the glandular regions. For gland segmentation, the authors have employed an active contour model which comes with certain limitations: long run time, proper parameters selection and initialization of contour. In another study [14], Olgun et al. formulated a novel texture descriptor specifically for adenocarcinoma to classify the images into normal, low grade and high grade. Their proposed approach is inspired by the local binary pattern method.

The gland segmentation is a pre-requisite step in all the morphological based approaches as well as in region-specific feature-based approaches. In all the above mentioned previous studies except [10], hand-crafted features and active contours have been used for gland segmentation. Recently there has been an increased interest of the research community in applying deep learning to the histology images for various tasks including segmentation and classification. In 2015, MICCAI GlaS challenge on segmentation of glands in H&E stained histology images was organised. In this challenge, out of 6 top ranking methods,

5 were based on CNNs [15]. In [10] and [16], UNET and multi-scale UNET are used for gland segmentation respectively. To improve the gland segmentation performance, there has been an encouraging number of architectures proposed in literature quite recently. CNNs have been extensively used for the classification of histology images as well. However, there is not much work on the analysis of adenocarcinoma. In [17], authors have employed a simple CNN to classify colorectal tissue images. Apart from adenocarcinoma, they have been extensively employed to other cancer types. In most of the previous work, GoogleNet and ResNet have been observed to be the utmost choice of the researchers due to their state-of-the-art results.

All the studies mentioned above on the classification of adenocarcinoma except [11] and [17] are focused on discriminating the malignant images from the normal/benign images and advancing it further to identify the extent of malignancy (low or high grade). While there are very few studies on identifying the non-cancerous abnormalities (hyperplasia, low-grade and high-grade dysplasia) which can lead to carcinoma at some point if left untreated. These abnormalities have been taken into account in our study and the proposed approach is designed while keeping these samples in mind. The morphological features of the glands have been shown to differentiate malignant glands from the normal glands with a high margin. But this may not be the case with abnormal non-cancerous glands since the abnormal glands may appear in circular/ellipsoidal shape and are of the same size as that of normal glands. Therefore, the morphological features of glands may not serve as a good representation of the pre-cancerous abnormalities. Contrary to morphological features, pixel level features or features that capture the internal structure of gland may perform better since the internal structure of abnormal glands appear to be different from the normal glands and is the focus of this study.

3. Clinical Aspect of Our Dataset

Apart from determining the grade of cancer, the identification of abnormalities is carried out in pathological practice. Some of the abnormalities can progress to develop into carcinoma depending on their extent of aberrance from normalities. Different treatment plans are advised for patients with different abnormalities. Therefore, correct identification is significantly important. For instance, small hyperplastic polyps (HP) are considered benign, most likely to be cured by simple excision and are unlikely to recur. Whereas, dysplastic polyps are irreversible, possess the ability to progress to carcinoma and the patient is subjected to continuous follow-up. Dysplasia is one of the critical precancerous class, particularly in case of inflammatory bowel disease since there is a possibility of progression from no dysplasia to low-grade dysplasia which may progress to high-grade dysplasia and then to carcinoma. However, in clinical practice, carcinoma often appears to form without following this pattern. Keeping this in mind, we have built a dataset consisting of images of normal, carcinoma (CA) and two types of abnormal (non-cancerous) tissues: hyperplastic polyp (HP) and tubular adenoma with low-grade dysplasia (TA_LG). The visual appearance of the glandular structure belonging to these classes is dissimilar such as 1) the normal glands appear circular or elliptic in shape, depending on the cutting plane, with well defined internal structure, 2) in hyperplastic polyp, there will be well-formed, elongated glands and crypts with serrated (saw tooth) or star-shaped lumen inside, 3) in tubular adenoma with low-grade dysplasia, the gland lining surrounded by the epithelial cells would not follow its normal spatial behaviour and these cells may change in size and their number and 4) in carcinoma, individual glands may not be differentiable, the extent of which depends on the grade of cancer. Example images of all these four classes obtained using our microscope-based image acquisition system are shown in Figure 1.

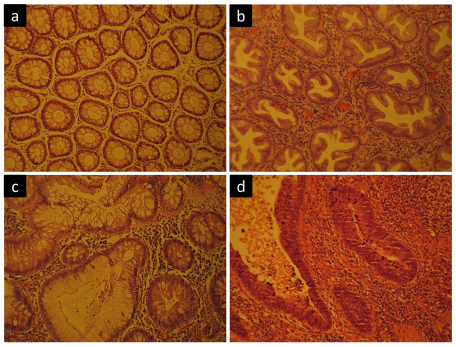

Figure 1: Examples images of four different classes of colorectal tissue: a) Normal, b) Hyperplastic polyp (HP), c) Tubular adenoma with low grade dysplasia (TA_LG) and d) Carcinoma (CA). These images are captured using RGB camera mounted over a microscope with halogen illumination.

4. Methodology

To reduce the correlation in the representation of colorectal images belonging to our different classes, we removed the non-glandular area from the images. To achieve this, we carried out a gland segmentation using a convolutional neural network (CNN) trained on a large dataset captured using a commercial whole slide image (WSI) scanner, different from the dataset designed for this study. To imitate the colour distribution of training images of this network, we pre-processed our dataset. Following the segmentation, we have adopted CNN based classification to learn the features of glandular structures. The overall flow of our methodology is shown in Figure 2.

4.1. Image Pre-processing

It is a general practice in computational pathology to have stain normalization step in a pipeline or to follow a stain invariant approach by performing stain augmentation. Due to the use of a microscope with halogen illumination, the appearance of our acquired dataset varies significantly from the images scanned using commercial WSI scanners. To scale down the stain difference between our acquired dataset and the dataset used to train the segmentation network, we performed stain normalization using Reinhard method [18].

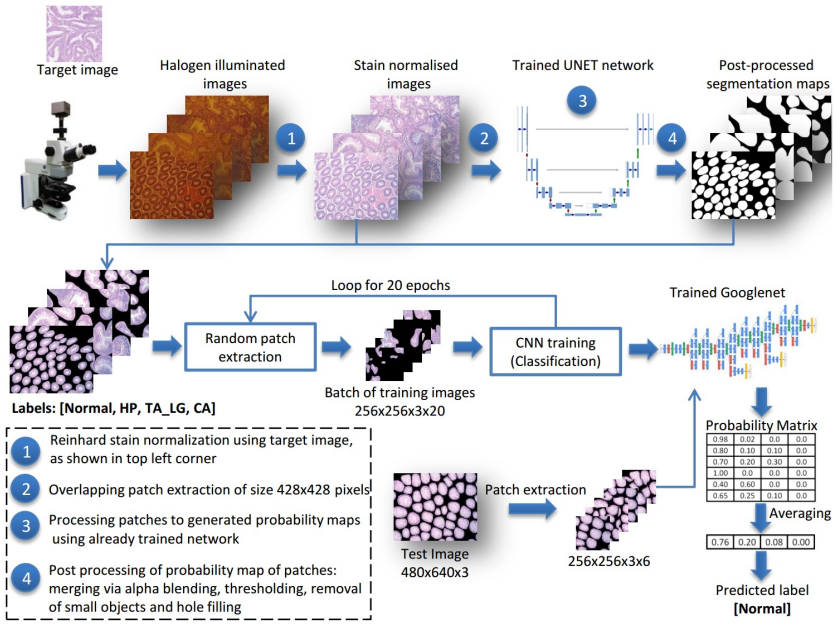

Figure 2: The overall system workflow.

4.2. Gland Segmentation

Gland segmentation was performed using CNN based UNET architecture proposed by Ronneberger et al. [19]. In a previous study [10], this network was adapted and trained on a very large colorectal dataset of RGB images of size 428×428 pixels extracted from Haematoxylin and Eosin (H&E) stained WSIs at 20x magnification. For further details on this network architecture and its training, readers are referred to [10]. Since our dataset was acquired at $10\times$ magnification, therefore to use this network, we resampled our dataset using a bicubic interpolation method at a scaling factor of 2. During inference, we extracted the overlapping patches of size 428×428 pixels because the network produces an output probability map of size 244×244 pixels, smaller than the input image size due to the unpadded convolution layers. The segmentation map for any given image was generated by merging these small-sized probability maps. After merging, we observed some artefacts around the border of merged maps. To reduce their appearance, we performed inference with patches

extracted with 25% more overlap. The fusion of the output probability maps was carried out using alpha blending. Thresholding was then performed on the fused segmentation maps to reduce the false positives and the threshold value was selected empirically. Moreover, hole filling was performed and to further reduce the appearance of false objects, area-based thresholding was applied.

4.3. Image Classification

Using the segmentation map, all the pixel values of the stain normalized images were set to zero if the corresponding pixels in the segmentation map were classified as background by the segmentation network. The resulting images were then used to train the CNN network for classifying the images into four classes. We employed three different networks: GoogleNet, ResNet50 and DenseNet for the classification of gland segmented RGB images. These networks have given outstanding results in patch-based classification of histopathology images. DenseNet, due to its connectivity pattern, requires less number of parameters as compared to ResNet and GoogleNet. However, all these networks with millions of parameters are considered deeper and are prone to over-fitting when trained with a small size dataset. To avoid network over-fitting to the training set, we increased the number of training images using different augmentation methods including flip, rotation and elastic distortion (both barrel and pincushion distortion). Augmentation was performed on the overlapping patches of size 428×428 pixels, extracted from images of size 480×640 pixels. With augmentation, we were able to generate more than 9000 patches per class. During the learning phase, the training patches of size 256×256 pixels were randomly cropped from the augmented images to further mitigate the chances of over-fitting. We performed inference over patches of size 256×256 pixels in a sliding window fashion across the test images. To evaluate our classification results against the ground truth labels assigned by the pathologist, instead of taking the final prediction from the network we took softmax probabilities for all the patches extracted from a test image. These probabilities were then averaged across each class and the maximum of averaged probabilities was used to

obtain the final prediction for a test image.

5. Experimental Material, Results and Discussion

5.1. Dataset

This study was reviewed and approved by the Qatar University's Institutional Review Board (QU-IRB) and Al-Ahli Hospital Ethical committee which includes Medical Director and members from the different departments (Internal Medicine, General Surgeon and Ophthalmology) of Al-Ahli Hospital before the study began. The colorectal tissue slides and the related clinical data were obtained from the Pathology and Laboratory Medicine lab at Al-Ahli hospital, Qatar after de-identification. These slides belong to 151 different patients and the informed patient consent was obtained from all subjects. A total of 164 tissue slides stained with H&E stain were obtained, taken from the year 2007 to 2016. For 141 patients, one biopsy slide per patient was provided while for the remaining 10 patients a maximum of 3 slides per patient were used. On each biopsy slide, the region for each class (Normal, HP, TA_LG and CA) was marked by the pathologist. Digital images of size 480×640 pixels were captured at 10x microscopic magnification from the marked regions using our visual field imaging system consisting of a Canon PowerShot A650IS digital camera mounted over a halogen illuminated microscope. For each class, 50 images were captured and hence in total 200 images were used for the experiments presented in this paper. Although, this is a small dataset in terms of a number of images available for training and validation. However, this dataset contains enough histological diversity since it is acquired from 151 different patients tissue samples. Therefore, it could be used to demonstrate the generalizability of any classification based method after increasing this dataset using augmentation.

5.2. Evaluation Measures

Since the network is trained to differentiate among different classes based on the glandular features, its performance greatly relies on the output of the

segmentation network. To evaluate the efficacy of the trained segmentation network on our dataset, we evaluated its output after thresholding against the ground truth gland boundary. For the evaluation, we calculate the segmentation accuracy at both the pixel and object levels. Two metrics are used for both cases: Dice index and Jaccard index. These indices measure the overlap between ground truth and network segmented binary images. At the object level, these two indices measure how well each network segmented gland overlaps with the ground truth gland and vice versa. For classification evaluation, commonly used measures are used: accuracy, sensitivity, specificity and F1-score. The formulation of pixel-level and object-level dice and Jaccard index is given in Supplementary Materials document.

5.3. Implementation Details

The gland segmentation masks were generated for all the images using adapted UNET architecture. For classification, various CNNs were trained using gland segmented images and ground truth labels provided by the pathologist. For validation purpose, each network was trained three times for cross-validation, with completely non-overlapping training and testing dataset. We used strong cross-validation where training and testing split was carried out at the patient level rather than at image level. To avoid overfitting and to compensate for the small dataset, we randomly extracted patches of size $256 \times 256 \times 3$ pixels from the images of size $480 \times 640 \times 3$ pixels in each epoch. All the networks were trained for 20 epochs with a batch size of 30 images and with the same initial weights and learning rate policies. The best network was selected to generate the results, based on the validation accuracy. For efficient back-propagation, we updated the parameters after every 10 iterations of batch processing (batch accumulation [20]). Stochastic gradient descent was used as an optimization method with a learning rate initially set at 0.01 and was decremented thrice by a factor of 10.

Table 1: Evaluation of segmentation maps against the ground truth with respect to each class.

	Normal	HP	TA_LG	CA	Average
$Dice_{pixel}$	0.904	0.820	0.844	0.748	0.829
$Jacc_{pixel}$	0.847	0.713	0.666	0.619	0.711
$Dice_{object}$	0.827	0.709	0.739	0.612	0.722
$Jacc_{object}$	0.779	0.615	0.552	0.501	0.612

5.4. Gland Segmentation Results

Table 1 shows the segmentation accuracy in terms of pixel-level and object-level Dice and Jaccard indices for the images of all four classes. We observed pixel-level index values in the range of 0.62 to 0.90, with the best overlap achieved for Normal images and the lowest overlap for the images of CA class which are challenging to segment due to a wide variation in their shape and appearance. Similar findings are observed in the case of object-level indices, thus demonstrating the performance of the network in identifying each gland as a separate object. These values are in the range of 0.50 to 0.87. The qualitative results of the segmentation maps generated for each class along with the ground truth boundary-filled cavities are shown in Figure 3.

5.5. Image Classification

The evaluation of our classification method is performed for the representative features extracted from three categories of images: a) the raw images I_{raw} using classical approach, b) the ground truth segmented images I_{GTS} and c) the network segmented images I_{NS}. The classification accuracy obtained with GoogleNet, ResNet and DenseNet for all three categories (I_{raw}, I_{GTS} and I_{NS}) are shown in Table 2. The accuracy obtained with these networks for each class is shown in Supplementary Fig. 1. The lowest, middle and highest-ranking categories are highlighted with green, blue and red colour respectively. GoogleNet achieved the best accuracy in each category and its per class F1-score can be found in Table 3. These results show that our segmentation guided approach surpassed the classical unsegmented approach. The average time for end-to-end

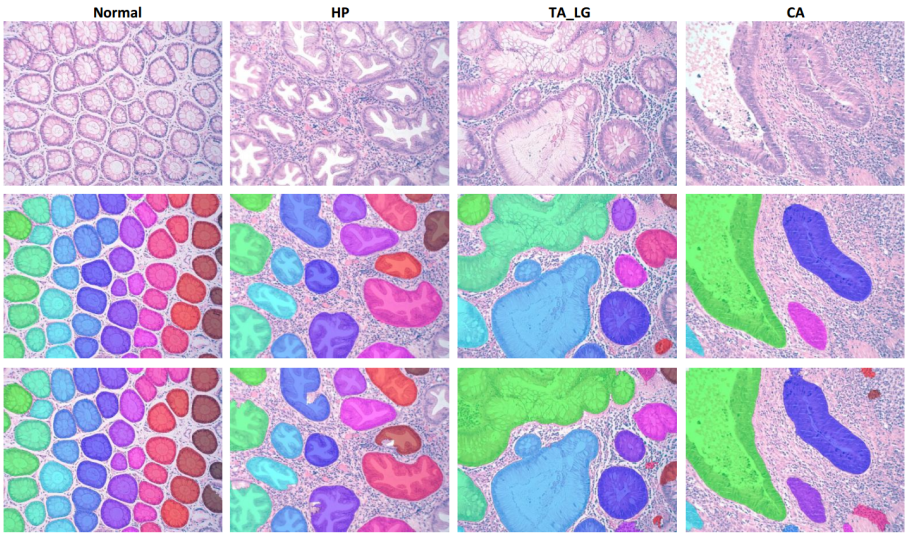

Figure 3: Qualitative results of gland segmentation. First, second and third row shows the original images, overlay of ground truth and UNET generated masks over the original images respectively. Each individually segmented gland is highlighted with a different color.

processing (both segmentation and classification) of an image of size 480×640 pixels with GoogleNet, DenseNet and ResNet50 are 1.35, 1.50 and 1.40 seconds respectively.

Table 4 shows the overall accuracy, sensitivity, specificity and F1-score along with two different categories. In the first category, CA class is considered as a positive class in which case the performance of the network segmented classification approach is comparable to that of ground truth segmented and the classical approach (using I_{raw} images). In the second category, TA_LG due to its increased risk of developing carcinoma is considered as a positive class along with CA. The results of this category and the overall results demonstrate the efficacy of segmentation guided approach in identifying the most significant classes when compared with the classical approach. Our comparison results obtained for the network generated and ground truth segmented images based on the above two categories show that the network can differentiate among classes using partial representative features obtained from glandular regions. This shows

that our network segmented system can identify these two positive classes, even when the segmentation of ROI is not very accurate. Furthermore, the prediction analysis of unsegmented and segmentation guided approach with the given dataset demonstrates the complexity of HP class as it is often misclassified as Normal and vice versa and has been observed to be confused with CA in light of our experimental results. This is demonstrated by the confusion matrices as shown in Supplementary Table I, II and III. It can be seen that none of the normal images is misclassified as CA and vice versa; the complexity of the given classification task is due to the addition of HP and TA_LG classes. We also conducted experiments to differentiate among different grades of cancer; the details of which are presented in the Supplementary Materials document.

5.5.1. Correlation between Segmentation and Classification

Our results demonstrated the efficacy of gland-guided approach when compared with the unsegmented approach. However, on comparing the results of the networks trained with I_{GTS} and I_{NS}, the former network performed better. This is due to the partial or over-segmentation of the glands and these findings can be demonstrated with Supplementary Fig. 2. These results are obtained by predicting the processed I_{GTS} using a network trained with I_{GTS}. The processed I_{GTS} were generated by using the ground truth masks that have been eroded or dilated with the rolling-ball structuring element of different window sizes. These results demonstrate that the network's classification performance declines when the glandular structures are partial or over-segmented. We also present our gland-guided classification result with three segmentation networks: FCN, SegNet and PspNet. The details of their architecture, training and performance are presented in the Supplementary Material document.

5.5.2. Visual Analysis of GoogleNet Features

The features learned by the GoogleNet for raw images and network segmented images at different layers of the network are shown in Figure 4. Feature maps from low-level (convolution layers before first inception module, C1 and

Table 2: Classification results obtained using different features for three different classification settings: I_{raw}, I_{GTS} and I_{NS}.

Method	I_{raw}	I_{GTS}	I_{NS}
GoogleNet	$\mathbf{0.830 \pm 0.02}$	$\mathbf{0.870 \pm 0.02}$	$\mathbf{0.850 \pm 0.03}$
DenseNet	0.794 ± 0.08	0.856 ± 0.06	0.819 ± 0.04
ResNet50	0.815 ± 0.01	0.845 ± 0.01	0.840 ± 0.00
LBP [21]	0.718 ± 0.07	0.835 ± 0.05	0.820 ± 0.02
LPQ [3]	0.752 ± 0.06	0.858 ± 0.04	0.812 ± 0.02
BSIF	0.728 ± 0.06	0.836 ± 0.05	0.791 ± 0.04
Morph-Features	-	0.640 ± 0.06	0.665 ± 0.01
BAM [10]	-	0.685 ± 0.16	0.600 ± 0.04

Table 3: F1-score obtained using GoogleNet.

	Normal	HP	TA_LG	CA	Overall
I_{raw}	0.793 ± 0.05	0.750 ± 0.06	0.864 ± 0.03	0.911 ± 0.03	0.830 ± 0.02
I_{GTS}	0.869 ± 0.09	0.806 ± 0.04	0.886 ± 0.05	0.922 ± 0.06	0.870 ± 0.02
I_{NS}	0.818 ± 0.12	0.765 ± 0.07	0.901 ± 0.06	0.913 ± 0.03	0.850 ± 0.03

C2), middle-level (second inception module, I2) and high-level layers (seventh inception module, I7) are observed. Feature maps shown in the first and second column demonstrate that the trained network has learned the features from both glandular and non-glandular tissue regions, specifically lower and middle layers of the network while the higher layer has given discriminatory weighting to most of the glandular region. In contrast, features maps shown in the third and fourth column, the network was explicitly restricted to learn from glandular tissue region only, which is the most distinguishable region of our dataset.

5.6. Texture Based Features

Texture features are shown to be the efficient representation of the histology images. Therefore, we also conducted experiments with texture features to analyse the effect of glandular guided based classification. We carried out experiments with three different texture features: local binary pattern (LBP)

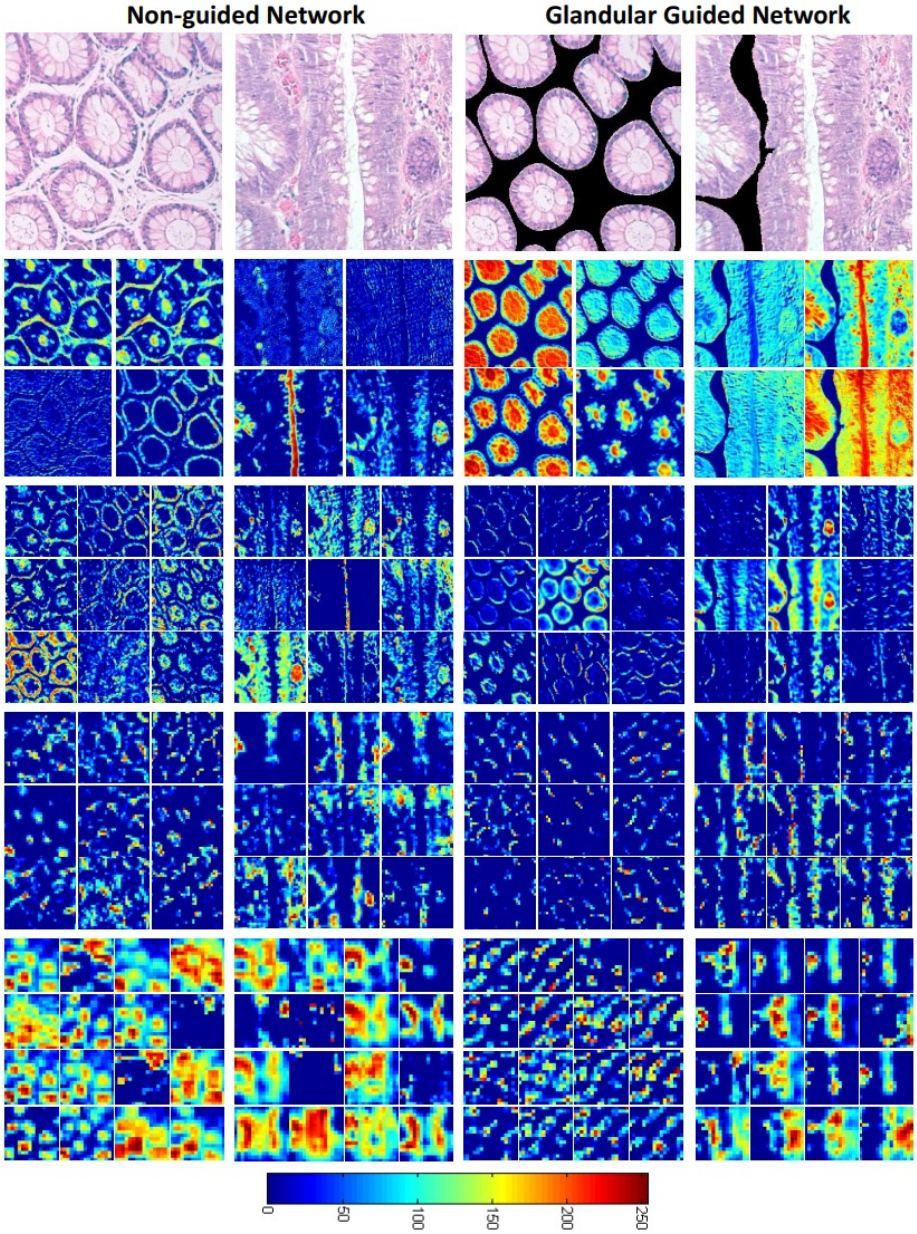

Figure 4: Output of various intermediate layers of non-glandular and glandular guided CNN. First row shows the input images while second, third, fourth and fifth rows show features maps obtained from GoogLeNet layers: *C1*, *C2*, *I2* and *I7* respectively. The scale bar shows the degree of activation.

Table 4: Comparison results of GoogleNet trained with three different settings.

Positive Class	Image Type	Accuracy	Specificity	Sensitivity	F1-score
CA	I_{raw}	0.955 ± 0.02	0.967 ± 0.01	0.920 ± 0.03	0.911 ± 0.03
	I_{GTS}	0.960 ± 0.03	0.974 ± 0.03	0.920 ± 0.03	0.922 ± 0.06
	I_{NS}	0.955 ± 0.02	0.960 ± 0.02	0.940 ± 0.02	0.913 ± 0.03
TA_LG + CA	I_{raw}	0.920 ± 0.02	0.929 ± 0.05	0.911 ± 0.08	0.918 ± 0.03
	I_{GTS}	0.931 ± 0.05	0.911 ± 0.08	0.951 ± 0.06	0.933 ± 0.05
	I_{NS}	0.945 ± 0.02	0.930 ± 0.05	0.961 ± 0.05	0.946 ± 0.02
Overall	I_{raw}	0.915 ± 0.01	0.943 ± 0.01	0.830 ± 0.02	0.829 ± 0.02
	I_{GTS}	0.935 ± 0.01	0.957 ± 0.01	0.870 ± 0.02	0.871 ± 0.02
	I_{NS}	0.925 ± 0.01	0.950 ± 0.01	0.850 ± 0.03	0.849 ± 0.03

[22], local phase quantization (LPQ) [23] and binarized statistical image features (BSIF) [24]. LBP and LPQ texture features have been presented as an effective representation of colorectal histology images for classification [21, 4, 3]. While BSIF, to the best of our knowledge, is used for the first time for representing colorectal tissue images. It is inspired by LBP and LPQ have been shown to perform better than LBP and LPQ for many non-medical dataset [24]. All these three features are evaluated in a classification setting by feeding them as an input to the SVM classifier. We experimented with four different kernels of SVM classifier: linear, polynomial, radial basis function (rbf) and sigmoid with fine-tuned parameters. From our experiments, we found that the rbf kernel for SVM gave the best results. The comparison result of these SVM kernels is presented in Supplementary Fig. 3. The analysis of texture features is presented in two perspectives: 1) how effective these features are in terms of classification accuracy when compared with features learned from the data itself (using CNNs) and 2) can texture features of glandular region serve as a fine representation for classification in comparison to texture features of non-segmented images. The quantitative analysis of these two perspectives is given in Table 2 and Supplementary Fig. 4. These results demonstrate that the features learned from the data itself perform better than the hand-crafted features, independently if these features are learned from either the gland seg-

mented images or non-segmented images. Similar to CNN, texture features extracted from the segmented glands outperform the features extracted from the non-segmented images. However, unlike CNN, the performance gap between the two (segmented vs non-segmented) is significant (5% to 10%) for all three texture features; which in case of CNN was observed to be 2% to 7%. The obtained results confirm the ability of gland segmented texture features to accurately classify the images.

5.7. Morphological Features

Additionally, we experimented with a different combination of morphological features of the segmented glands. The classification was performed using all possible combinations of the number of glands in an image with ten different morphological features of glands: area, convexity, roundness, aspect ratio, elongation, solidity, form factor, compactness, eccentricity and extent. The best accuracy was obtained with a combination of the number of glands, roundness, solidity and area of the gland and is observed to be less than the accuracy obtained using CNN and texture descriptors. We also evaluated our approach with one published study [10] for which the implementation was available to us. This study quantifies the morphology of the glandular structures using a BAM measure which has been shown to perform well for colorectal grading since the shape of gland varies across different grades. However, the pre-cancerous glands tend to retain their spherical shape; therefore, BAM and other morphological features have failed to perform well for the given task. The accuracy obtained using these morphological features is shown in Table 2. The values reported in Table 2, 3 and 4 are *mean* ± *standard deviation*.

6. Conclusion

The histology images belonging to two different classes may have similar tissue sub-regions which reduces the boundary distance between the classes. Therefore the performance of a classifier is hampered. The experimental results

of this study demonstrate that the classification performance improves if the model learns from the distinctive features in a histology image. To deal with this problem and the complexity of pre-cancerous class in the colorectal dataset, we confined the CNN to learn features from the segmented glands which can be used as a fine representation for each class of our interest. In addition to CNN based learning, we also evaluated our framework using some handcrafted texture features. We also got the same finding from our experimental results with handcrafted features. From the variety of features used in this study, we infer that our glandular guided method performs better classification irrespective of any type of features (either data learned or handcrafted).

References

[1] B. Vogelstein, E. R. Fearon, S. R. Hamilton, S. E. Kern, A. C. Preisinger, M. Leppert, A. M. Smits, J. L. Bos, Genetic alterations during colorectal-tumor development, New England Journal of Medicine 319 (9) (1988) 525–532.

[2] F. T. Bosman, F. Carneiro, R. H. Hruban, N. D. Theise, WHO classification of tumours of the digestive system., no. Ed. 4, World Health Organization, 2010.

[3] S. Kunhoth, S. Al Maadeed, Multispectral biopsy image based colorectal tumor grader, in: Annual Conference on Medical Image Understanding and Analysis, Springer, 2017, pp. 330–341.

[4] R. Peyret, A. Bouridane, F. Khelifi, M. A. Tahir, S. Al-Maadeed, Automatic classification of colorectal and prostatic histologic tumor images using multiscale multispectral local binary pattern texture features and stacked generalization, Neurocomputing (2017).

[5] R. Farjam, H. Soltanian-Zadeh, K. Jafari-Khouzani, R. A. Zoroofi, An image analysis approach for automatic malignancy determination of prostate

pathological images, Cytometry Part B: Clinical Cytometry 72 (4) (2007) 227–240.

[6] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, J. Tomaszewski, Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology, in: Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on, IEEE, 2008, pp. 284–287.

[7] K. Nguyen, A. K. Jain, R. L. Allen, Automated gland segmentation and classification for gleason grading of prostate tissue images, in: Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE, 2010, pp. 1497–1500.

[8] Y. Peng, Y. Jiang, L. Eisengart, M. A. Healy, F. H. Straus, X. J. Yang, Computer-aided identification of prostatic adenocarcinoma: Segmentation of glandular structures, Journal of pathology informatics 2 (2011).

[9] T. Gultekin, C. F. Koyuncu, C. Sokmensuer, C. Gunduz-Demir, Two-tier tissue decomposition for histopathological image representation and classification, IEEE transactions on medical imaging 34 (1) (2015) 275–283.

[10] R. Awan, K. Sirinukunwattana, D. Epstein, S. Jefferyes, U. Qidwai, Z. Aftab, I. Mujeeb, D. Snead, N. Rajpoot, Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images, Scientific reports 7 (1) (2017) 16852.

[11] J. N. Kather, C.-A. Weis, F. Bianconi, S. M. Melchers, L. R. Schad, T. Gaiser, A. Marx, F. G. Zöllner, Multi-class texture analysis in colorectal cancer histology, Scientific reports 6 (2016) 27988.

[12] R. Awan, S. Al-Maadeed, R. Al-Saady, Using spectral imaging for the analysis of abnormalities for colorectal cancer: When is it helpful?, PloS one 13 (6) (2018) e0197431.

[13] A. Chaddad, C. Desrosiers, A. Bouridane, M. Toews, L. Hassan, C. Tanougast, Multi texture analysis of colorectal cancer continuum using multispectral imagery, PloS one 11 (2) (2016) e0149893.

[14] G. Olgun, C. Sokmensuer, C. Gunduz-Demir, Local object patterns for the representation and classification of colon tissue images, IEEE journal of biomedical and health informatics 18 (4) (2014) 1390–1396.

[15] K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, A. Bhm, O. Ronneberger, B. B. Cheikh, D. Racoceanu, P. Kainzi, M. Pfeiffer, M. Urschler, D. R. Snead, N. M. Rajpoot, Gland segmentation in colon histology images: The glas challenge contest, Medical image analysis 35 (2017) 489–502.

[16] J. Li, K. V. Sarma, K. C. Ho, A. Gertych, B. S. Knudsen, C. W. Arnold, A multi-scale u-net for semantic segmentation of histological images from radical prostatectomies, in: AMIA Annual Symposium Proceedings, Vol. 2017, American Medical Informatics Association, 2017, p. 1140.

[17] F. Ponzio, E. Macii, E. Ficarra, S. Di Cataldo, Colorectal cancer classification using deep convolutional networks-an experimental study., in: BIOIMAGING, 2018, pp. 58–66.

[18] E. Reinhard, M. Adhikhmin, B. Gooch, P. Shirley, Color transfer between images, IEEE Computer graphics and applications 21 (5) (2001) 34–41.

[19] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[20] Y. A. LeCun, L. Bottou, G. B. Orr, K.-R. Müller, Efficient backprop, in: Neural networks: Tricks of the trade, Springer, 2012, pp. 9–48.

[21] R. Peyret, A. Bouridane, S. A. Al-Maadeed, S. Kunhoth, F. Khelifi, Texture analysis for colorectal tumour biopsies using multispectral imagery, in:

Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, IEEE, 2015, pp. 7218–7221.

[22] T. Ahonen, J. Matas, C. He, M. Pietikäinen, Rotation invariant image description with local binary pattern histogram fourier features, Image analysis (2009) 61–70.

[23] V. Ojansivu, J. Heikkilä, Blur insensitive texture classification using local phase quantization, in: International conference on image and signal processing, Springer, 2008, pp. 236–243.

[24] J. Kannala, E. Rahtu, Bsif: Binarized statistical image features, in: Pattern Recognition (ICPR), 2012 21st International Conference on, IEEE, 2012, pp. 1363–1366.

Authors' Biographies

Ruqayya Awan holds a Master's degree in Computer Science from National University of Science and Technology, Pakistan. She is currently a PhD student based in the Tissue Image Analytics lab within the Computer Science Department, University of Warwick, UK. She is working on the analysis of pathology and cytology images and her research is funded by the Warwick Chancellors International Scholarship.

Somaya Al-Maadeed is an Associate Professor of Computer Science and heading the Computer Science and Engineering Department at Qatar University. She holds a PhD degree from Nottingham University, with interests in pattern recognition, computer vision, deep learning and character recognition. She is a coordinator of the Computer Vision and Robotics Research Group and enjoys excellent collaboration with industry and international institutions.

Rafif Al-Saady is a consultant histopathologist and cytopathologist in a Pathology and Laboratory Medicine Department of Al-Ahli Hospital, Doha. Her

expertise lies in preparing and reporting cytology and biopsy specimen, developing and organizing programs for cancer control and teaching of pathology to under and postgraduate students. She has a number of publications in medical journals.

Ahmed Bouridane is a full Professor in Machine Intelligence and leads the Computational Intelligence and Visual Computing Group at Northumbria University (UK). His research interests are in imaging for forensics and security, biomedical engineering and cybersecurity. He has published more than 350 publications and two research books on imaging for forensics and security. He is a Senior Member of IEEE.