

Multispectral Band Selection and Spatial Characterization: Application to Mitosis Detection in Breast Cancer Histopathology

H. Irshad^{a,c,*}, A. Gouaillard^{d,e}, L. Roux^{a,c}, D. Racoceanu^{b,c}

^aUniversity Joseph Fourier, Grenoble 1, France - IPAL UMI 2955

^bSorbonne Universités, UPMC Univ Paris 06, France - IPAL UMI 2955

^cFrench National Center for Scientific Research, IPAL UMI 2955 Singapore

^dCoSMo Software, Boston, MA, USA

^eTemasys Communications, Singapore

Abstract

Breast cancer is the second most frequent cancer. The reference process for breast cancer prognosis is Nottingham grading system. According to this system, mitosis detection is one of three important criteria required for grading process and quantifying the locality and prognosis of a tumor. Multispectral imaging, as relatively new to the field of histopathology, has the advantage, over traditional RGB imaging, to capture spectrally resolved information at specific frequencies, across the electromagnetic spectrum. This study aims at evaluating the accuracy of mitosis detection on histopathological multispectral images. The proposed framework includes: selection of spectral bands and focal planes, detection of candidate mitotic regions and computation of morphological and multispectral statistical features. A state-of-the-art of the methods for mitosis classification is also provided. This framework has been evaluated on MITOS multispectral dataset and achieved higher detection rate (67.35%) and F-Measure (63.74%) than the best MITOS contest results [46]. Our results indicate that selected multispectral bands have more discriminant information than a single spectral band or all spectral bands for mitotic figures, validating the interest of using multispectral images to improve the quality of the diagnostic in histopathology.

Keywords: histopathology, breast cancer, multispectral images, spectral

*Corresponding author, E-mail: humayun.irshad@gmail.com, Tel. No. +65-64082601

bands selection, texture characterization, classification, object detection, features extraction

1. Introduction

According to the International Agency for Research on Cancer (IARC) in its GLOBOCAN 2012 estimated cancer incidence, mortality and prevalence worldwide [1], breast cancer is the second most commonly diagnosed cancer worldwide after lung cancer. In 2012, there was an estimated 1.677 million new breast cancer cases and 522,000 deaths from breast cancer.

The reference process for breast cancer prognosis, recommended by the World Health Organization, is histologic grading that combines tubule formation, nuclei atypia and mitotic count [2, 3]. This assessment of tissue sample is synthesized into a diagnosis that would help the clinician to determine the best course of therapy. Several computer aided diagnostic solutions exist for the detection of tubule formation [4, 5] and nuclei atypia [6, 7, 8, 9] but only a few are dedicated to mitosis detection [10, 11].

In histopathology, hematoxylin and eosin (H&E) is a well-established staining technique, exploiting intensity of stains in the tissue images to quantify the nuclei and other structures related to cancer developments [13]. In this context, image processing techniques are devoted to accurate and objective quantification and localization of cancer evolution in specific regions of the tissue such as cytoplasm, membranes and nuclei [14]. From the chromatic viewpoint, nuclear regions are characterized by non-uniform stain intensity and color, thus preventing a trivial classification based on color separation. In addition, the superposition of tissue layers, as well as the diffusion of the dyes on the tissue surface, may bring the stains to contaminate the background or other cellular regions, which are different from their specific target.

One of the most difficult challenges in histopathological dataset analysis is spatial analysis, more specifically automated nuclei detection and classification [15, 12]. The objective of nuclei classification is to assign different labels to different types of nuclei as normal, cancer, mitotic, apoptosis, lymphocytes etc. In addition, quantitative characterization is important not only for clinical applications (e.g., to reduce/eliminate inter- and intra-observer variation in diagnosis) but also for research applications (e.g., to understand the biological mechanisms of the disease process) [16].

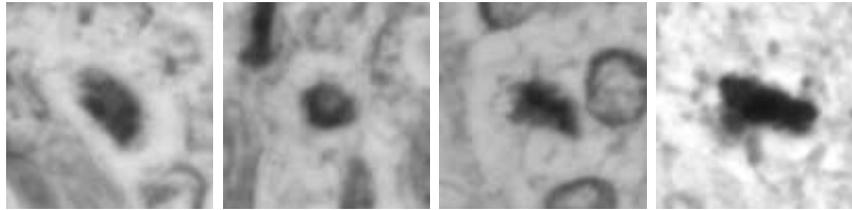
Image analysis in cytology has been studied for years and numerous solutions [17, 30, 31, 32, 33] have been proposed in the literature. The application

of these solutions to histopathology is rather complicated due to the radical differences between the two imaging modalities and to the highly complex image characteristics. Indeed, in the case of histology images, cellular structures and functions are studied embedded in the whole tissue structure, presenting various cells architecture (gland formation, DCIS), very difficult to handle with usual pattern recognition techniques. Nevertheless, recent works [34, 10, 11, 12] show great potential for computer assisted diagnostic of histopathological datasets for breast cancer grading.

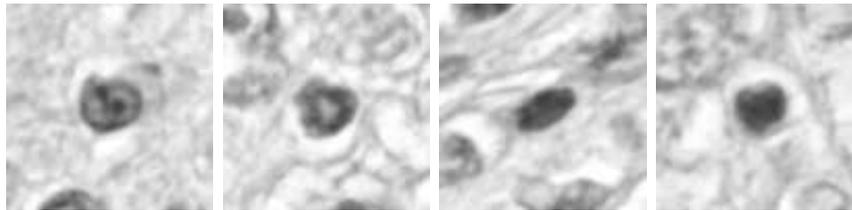
Multispectral imaging has the advantage over traditional RGB imaging to retrieve spectrally resolved information of a tissue image scene at specific frequencies across the electromagnetic spectrum. Multispectral imaging system captures images with accurate spectral content, correlated with spatial information, by revealing the chemical [29] and anatomic features of histopathology [35, 36]. This modality provides option to biologists and pathologists to see beyond the RGB image planes to which they are accustomed. Recent publications [37, 38, 39, 40] have begun to explore the use of extra information contained in such spectral data. Specifically, a comparison of spectral methodologies demonstrates the advantage of multispectral data [41, 42]. The added benefit of multispectral imaging for analysis in routine H&E stained histopathological images, however, is still largely unknown, although some promising results are presented in [43, 37, 40, 39]. As far as we know, there is no existing study of the use of multispectral imaging for automation of mitosis detection in breast cancer histopathology.

In our study, we propose a framework [11] for mitosis detection in breast cancer multispectral histopathology. This framework addresses two important questions: First, does the spatial-spectral analysis on selected spectral bands (as opposed to spatial analysis on single spectral or spatial-spectral analysis of all the spectral bands) suffice for efficient classification of mitotic and non-mitotic figures. An obvious advantage of using selected spectral bands is its reduced computational and storage complexity. Second, how effective are the multiple features for discrimination of mitotic and non-mitotic figures, as compared to one type of features? Some examples of mitotic and non-mitotic nuclei are shown in Figure 1.

The remainder of the paper is organized as follows. Section 2 reviews the state-of-the-art of multispectral methods, particularly in object or region detection in histopathology, related to this research work. Section 3 describes the dataset used for this study. Section 4 describes the proposed framework for mitosis detection in breast cancer multispectral histopathol-



(a) Mitotic nuclei.



(b) Non-mitotic nuclei.

Figure 1: Examples of mitotic and non-mitotic nuclei in breast cancer multispectral histopathology. It is difficult to manually differentiate between mitotic and non-mitotic nuclei due to similarity.

ogy. Experiments and results are presented in section 5. Section 6 contains the discussion part. Finally, the concluding remarks with future work are presented in section 7.

2. Literature Review

Multispectral imaging uses more than three spectral filters (not like RGB color images) to capture a series of images having spectral and spatial information. This spectral and spatial information has been used in remote sensing [18, 19, 20, 21], biometric systems [22, 23], plant and animal disease identification [24, 25, 36, 26, 27]. The main idea for extracting discriminative information from multispectral imaging is the use of combined spectral and spatial information for discrimination of regions or objects. In earlier works, Holmquist et al. [28] utilized spectrophotometric information for cervical cell segmentation in Papanicolaou-stained images. They found that two spectral bands were more appropriate for cervical cells segmentation as compared to single spectral band.

Fernandez et al. [37] coupled high-throughput Fourier transform infra-red spectroscopic imaging of tissue microarrays with statistical pattern recognition of spectra indicative of endogenous molecular composition for histopatho-

logical characterization of prostate tissue. They explicitly defined metrics consisting of spectral features that have a physical significance related to tissue biochemistry and facilitating the measurement of cell types.

We found few methods in the multispectral literature for spectral and spatial characterization of histopathological images. Some of them employed single spectral band of multispectral imaging [44, 34] and others used multiple spectral bands of multispectral imaging [40, 39, 45]. Some methods computed one type of features on single spectral band for quantitative analysis. Masood and Rajpoot [44] proposed a colon biopsy classification method based on spatial analysis of hyperspectral images. First, spectral band 588 nm was selected, as it is the one that seemed to contain more textural information. Then, using circular local binary pattern algorithm, spatial analysis of patterns was represented by a feature vector in the selected spectral band. Later, classification was achieved using subspace projection methods like principal component analysis (PCA), linear component analysis (LCA) and support vector machine (SVM).

Some methods computed different types of features on single spectral band for quantitative analysis. Malon and Cosatto [34] combined manually designed nuclear features with the learned features extracted by convolutional neural network for mitosis detection and achieved the best F-Measure (59%) on multispectral dataset during ICPR contest 2012 [46]. First, focal plane number five was selected as it was clearly focused. Second, two spectral bands were selected using PCA to extract the top two eigenvectors from a set of 10 spectral bands of H&E stained multispectral images. Third, two step thresholding was applied on first eigenvectors (hematoxylin image) to obtain candidate blobs. Fourth, a set of shape, contour, pixel and texture features was computed on the selected spectral band only. Fifth, log likelihoods of class membership were computed using convolutional neural network classifier for each patch of candidate blob. In the last, the SVM classifier was used to classify each blob as either mitotic or non-mitotic blob using output of convolutional neural network along with feature vector. This approach discards additional potentially relevant information from other spectral bands.

Instead of limiting themselves to a single spectral band, some authors use multiple and sometime even all spectral bands, from a given dataset. Boucheron et al. [45] presented a study in which the additional spectral bands have additional useful information for nuclear classification in histopathology as compared to the three standard bands of RGB imagery. Using all spectral bands, they reported a 0.79% improvement in performance compared to

the next best performing image type. Similarly, Wu et al. [39] proposed a multilayer conditional random field model using a combination of low-level cues and high-level contextual information for nuclei separation in high dimensional data set obtained through spectral microscopy. In this approach, the multilayer contextual information is extracted to interpret spectral data with dynamically imposed pairwise constraints along the neighboring spectral bands. It is an unsupervised process, which efficiently helps to suppress segmentation errors caused by intensity inhomogeneity and variable chromatin texture. Khelifi et al. [40] proposed a spatial and spectral gray level dependence method in order to extend the concept of gray level co-occurrence matrix by assuming the presence of texture joint information between spectral bands. Some spectral bands have more relevant information for specific object or region classification than others. This approach is limited to a single spatial feature computed from all spectral bands. Approaches like [45, 39, 40] used all available spectral bands but were limited to one type of features only. One possible improvement in object classification is multispectral analysis using more types of features. Another possibility of additional improvement is the selection of spectral bands by minimization of the redundancy and maximization of the relevancy.

In multispectral imaging, the question is how to select the correct spectral bands from the spectral range to best characterize the problem. More specifically, there are strong correlations between spectral bands and some spectral bands cannot discriminate between mitotic nuclei and others; this is the reason why the effective dimension of spectral bands for classification is less than the total number of spectral bands. In literature, we found one possible solution that uses information theory, more specifically, by means of measures based on the mutual information (MI) to feature selection for pixel classification. Furthermore, Martinez-Uso et al. [47] proposed a hierarchical clustering framework based on MI for spectral bands selection. Kamandar et al. [48] used minimum redundancy maximum relevance (mRMR) technique [49] for spectral bands selection in AVARIS data.

Each mitotic region has different level of relevant information in different spectral bands. In the proposed framework, we address the shortcomings of previous works, including (1) selection of spectral bands; (2) analysis of statistical-spectral features in selected spectral bands rather than single [44, 34] or all spectral bands [45, 39, 40] and (3) selection of the best classifier for discrimination of mitotic figures from other microscopic objects. The main novel contributions of the proposed work are:

1. Three different methods for spectral bands selection including relative spectral absorption of different tissue components, spectral absorption of H&E stains and mRMR technique.
2. Computation of a morphological & multispectral statistical features vector (MMSF) containing intensity, texture and morphological features which leverage discriminant information from a given candidate across selected spectral bands for classification of mitotic and non-mitotic figures.
3. An extensive investigation of classifiers and inference of the best one for mitotic figures classification.

3. Dataset

We evaluated the proposed framework on multispectral MITOS dataset [50] a freely available dataset of medical images where localization of mitotic figures have been provided by experienced pathologists. To the best of our knowledge, this is the only available de facto gold standard dataset of both multispectral and color images. It provides basis not only for comparison between our proposed framework and other previous frameworks, but also for comparison between multispectral and color images. Using this dataset, IPAL CNRS laboratory organized a contest during ICPR 2012 [46]. They also provided the performance metrics to evaluate the framework for mitosis detection. We used this dataset and compared our results against the results of contestants.

The data set is made up of 200 images coming from five different slides scanned at 40X magnification using a 10 spectral bands microscope. There are 40 images per slide and each image has a size of $251.6 \times 251.6 \mu\text{m}^2$ (that is an area of 0.063 mm^2). The 200 images contain a total 322 mitotic figures. The training data set consists of 140 images containing 224 mitotic figures and the evaluation data set consists of 60 images containing 98 mitotic figures [46].

The spectral bands are all in the visible spectrum. There is some spectrum overlapping for the spectral bands. In addition, for each spectral band, the digitization has been performed at 17 different focal planes (17 layers Z-stack), each focal plane being separated from the other by 500 nm. Therefore, for each image, there is a stack of 170 files (10 spectral bands and 17 focal planes for each spectral band). Fig. 2 shows the spectral coverage of each of the 10 spectral bands of the multispectral microscope.

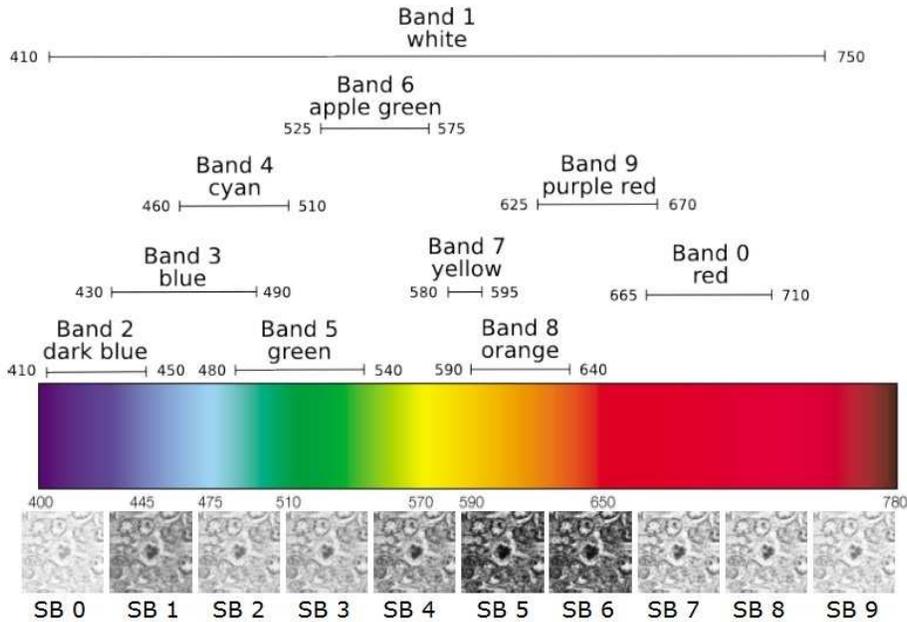


Figure 2: Spectral bands of the multispectral microscope and examples for each spectral band (SB).

4. Proposed Framework

In this paper, we propose a framework for mitosis detection in breast cancer multispectral histopathology as shown in Fig. 3. The proposed framework has five main steps. Step one performs a selection of the most informative focal plane based on maximum gradient information of mitotic figures from background. This is explained in Appendix A. Step two is responsible for the selection of relevant spectral bands for the objective of mitotic figures detection. Candidates for mitotic figures are detected in step three. Then, in step four, intensity and texture features across selected spectral bands are computed for each detected candidate. In addition, using segmented regions of detected candidates, morphological features are also computed and added to the feature vector. During step five, candidates are classified into mitosis and non-mitosis classes using decision tree (DT), multilayer perceptron (MLP) as well as linear and non-linear support vector machine (L- and NL-SVM) classifiers. A side advantage of performing the statistical analysis on multiple selected spectral bands simultaneously is to investigate whether im-

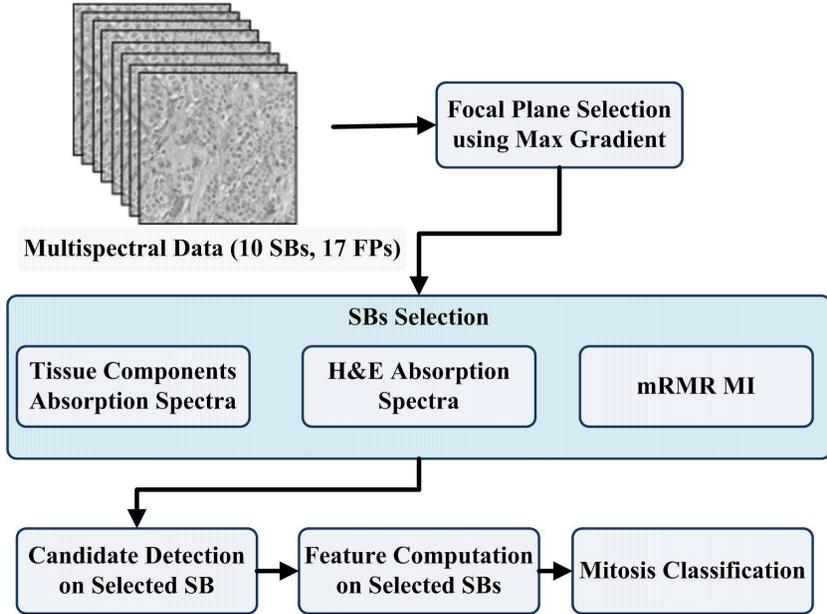


Figure 3: Proposed Framework (SBs: spectral bands, MI: Mutual Information).

provement in accuracy can be achieved with multispectral statistical features computation in selected spectral bands over those methods which use single spectral band [44, 34] or all spectral bands [45, 39, 40].

4.1. Spectral Bands Selection

4.1.1. Method 1: Tissue Spectral Absorption

The main tissue components visible in the data set images can globally be categorized into fat, stroma and epithelial nuclei as shown in Fig. 4. As our purpose is the detection of mitotic nuclei only, we further subdivided epithelial nuclei into mitotic and non-mitotic nuclei. We selected 200 image patches, a patch being a region of interest of 150×150 pixels, for each tissue components and computed the spectral absorption responses of each tissue components for the available 10 spectral bands as shown in Fig. 5. In 5(a), fat tissue is negligible as it has very low absorption response. Moreover, mitotic and non-mitotic nuclei contributions are indistinguishable. We select mitotic nuclei and stroma curves to compute the maximum differentiation in contribution to pixel intensity. In 5(b), one can see that bands 7, 8 and 9 exhibit the biggest difference between mitotic nuclei and stroma contributions. They are the best candidates for maximum differentiation.

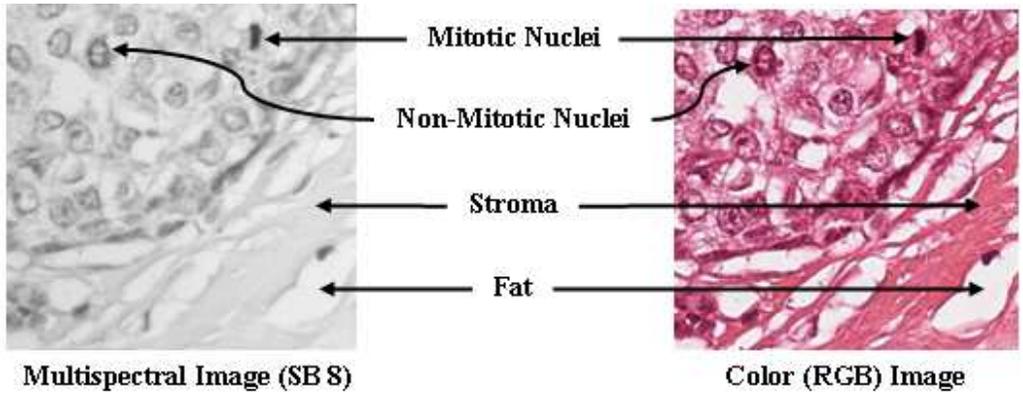
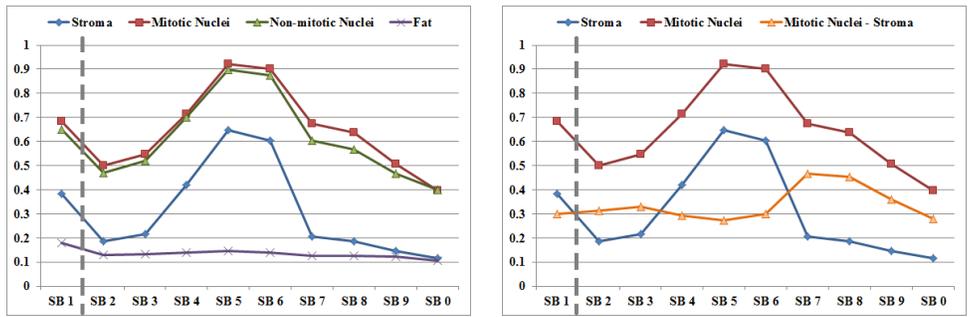


Figure 4: Example of different components of breast tissue in H&E stained histopathological image. Left image sample is taken from spectral band 8, focal plane 6 of multispectral microscope; right image is taken from Aperio Slide Scanner.



(a) Normalized absorption spectra of four tissue components in 10 SBs (b) Difference of mitotic nuclei and stroma absorption spectra in 10 SBs

Figure 5: Normalized absorption spectra of four tissue components in 10 spectral bands (SBs). Note that SB 1 (white band), in nature, is different from other SBs and may serve as reference as it covers the whole visible spectrum and contains all the information that other bands are containing, although at a lower resolution. It is separated from other SBs by a dotted line.

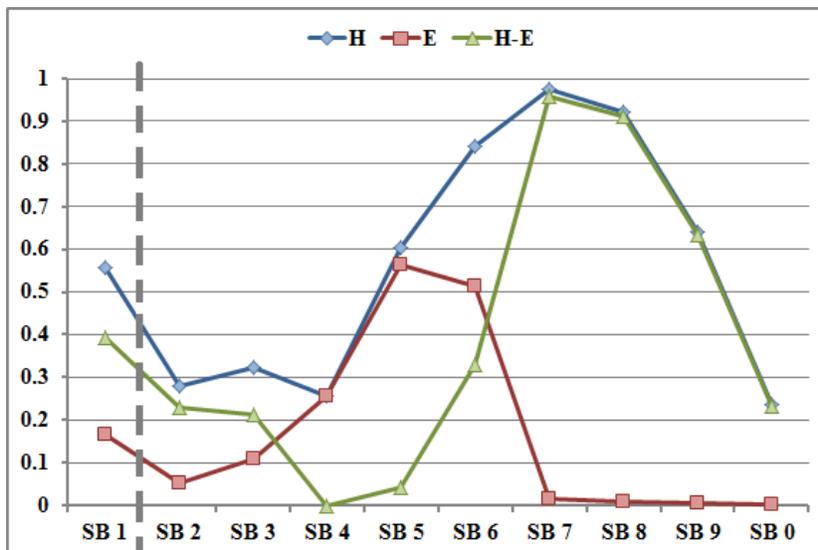


Figure 6: Normalized plot of hematoxylin (blue line) and eosin (red line) dyes absorption spectra in multispectral images and the difference of hematoxylin and eosin (green line).

4.1.2. Method 2: Hematoxylin and Eosin Spectral Absorption

In this method, we reconstruct the spectrum of a pixel by using staining characteristics of tissue components to select the optimal number of spectral bands for mitosis discrimination in H&E stained multispectral histopathology. To illustrate the possible correlation between spectral bands and the staining characteristics of the spectral samples, the plot of hematoxylin and eosin dyes spectral absorptions are shown in Fig. 6 (this plot is derived from the work of Bautista and Yagi [51]). Hematoxylin stains nuclei material, while eosin stains both nuclei and cytoplasm. The H–E plot in Fig. 6 shows the difference of absorption between hematoxylin and eosin. The spectral bands for which H–E is maximum are more suitable for discrimination between nuclei and cytoplasm. The difference is maximum in spectral bands 7 and 8 with almost zero eosin absorption. Therefore, these spectral bands should be good options for mitosis detection.

4.1.3. Method 3: mRMR Technique

In this method, mRMR technique [49] is used for selection of spectral bands. Selection is based on two criteria; minimum redundancy $R(S, c)$ and maximum relevance $D(S, c)$. The relevancy of selected spectral bands to class

labels has been measured by average of mutual information (MI) between each spectral band and class label. Their redundancy is measured by an average of MI between each pair of spectral bands. The average relevancy of selected spectral bands is defined as:

$$D = \frac{1}{|S|} \sum_{s_i \in S} \text{MI}(s_i; c_j) \quad (1)$$

where S denotes the selected spectral bands set, $|S|$ denotes the number of selected spectral bands, c_j denotes j^{th} class label in class set C , s_i denotes i^{th} spectral band in S and MI is mutual information between spectral band s_i and class label c_j . MI is computed using entropy as

$$\text{MI}(S; C) = H(S) - H(S|C) \quad (2)$$

where

$$H(S) = - \sum_{s_i \in S} p(s_i) \log_2(p(s_i)) \quad (3)$$

and

$$H(S|C) = - \sum_{s_i \in S} \sum_{c_j \in C} p(s_i, c_j) \log_2(p(s_i|c_j)) \quad (4)$$

are entropy functions that calculate the uncertainty of the spectral bands and the class labels. In equations (3) and (4), $p(s_i)$ is probability density function of s_i and $p(s_i|c_j)$ is conditional probability density function of s_i and c_j . By maximizing D for full spectral bands set S_T , we can select a spectral bands set S with maximum relevance for discrimination of mitotic candidates with observing spectral bands set S_T .

It is likely that selected spectral bands have rich redundancy. Therefore, the following minimum redundancy $R(S, c)$ is added to select mutually exclusive spectral bands:

$$R = \frac{1}{|S|^2} \sum_{s_i, s_j \in S} \text{MI}(s_i; s_j) \quad (5)$$

$\text{MI}(s_i; s_j)$ is maximum when two spectral bands s_i and s_j have functional dependency and $\text{MI}(s_i; s_j) = 0$ if s_i and s_j are statistically independent. By

Table 1: Spectral Bands (SBs) Mutual Information (MI) Measure.

SBs	MI	Accumulated MI	Accumulated MI%
SB 8	3.60	3.60	33%
SB 9	3.59	0.95	42%
SB 7	3.38	0.94	51%
SB 6	3.18	0.93	60%
SB 2	3.16	0.92	69%
SB 1	3.11	0.91	78%
SB 3	3.05	0.89	86%
SB 0	2.99	0.88	91%
SB 4	2.94	0.85	95%
SB 5	2.85	0.82	100%

minimizing R for selected spectral bands, we selected spectral bands set with minimum redundancy.

Selection of Spectral Bands in Set (S) in Equations (1), (2), (3), (4) and (5): The incremental search method was used to find the n spectral bands from the set $\{S_T - S_{n-1}\}$, maximizing the following condition expression:

$$\max_{s_i \in S_T - S_{(n-1)}} \left[\text{MI}(s_i; c) - \frac{1}{n-1} \sum_{s_j \in S_{(n-1)}} \text{MI}(s_i; s_j) \right] \quad (6)$$

The image patches, used in computation of spectral absorption of different tissue components (as described in section 4.1.1), were divided into two classes. The non-mitosis class consisted of three tissue components including fat, stroma and non-mitotic nuclei, and the remaining samples belonged to mitosis class. We performed mRMR on these image patches and their MI with ranking are shown in Table 1.

4.2. Candidate Detection on the Spectral Band Having Highest Mutual Information

First, we select the spectral band that has highest MI for candidate detection, that is band 8. Using this spectral band, we smoothed the input image with a median filter. Then, we performed binary thresholding on smoothed image using threshold T as shown in Figure 7(b):

$$I^B(x, y) = \begin{cases} 1, & \text{if } I(x, y) < T \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

The value of T is selected from the histogram of mitotic and stroma regions on spectral band 8 as shown in Appendix B (in Figure 12).

Morphological opening and closing are applied to the binary image $I^B(x, y)$ to merge the clustered regions into large regions, fill holes and eliminate too small regions. Then we segment the boundaries of candidates using active contour models with a level set implementation [59]. The key steps for the segmentation method are as follows:

1. A given nuclei contour $\varpi(t)$ is presented as the zero level of the signed distance function $\psi(x, t)$. Formally, $\varpi(t) = \{x : \psi(x, t)\}$
2. Later, active contour formation is used as:

$$\frac{\partial \psi}{\partial t} = f(I)(\alpha b + \beta k)|\nabla \psi| + \gamma \nabla f \cdot \nabla \psi \quad (8)$$

where α, β and γ are user-defined settings for the relative scaling of the terms, f refers to the image-based feature function that is minimized at nuclei boundary and remains high elsewhere, b is a balloon force that is added to evolve the curve outwards, k is the curvature along the normal to the level set contour and $\nabla f \cdot \nabla \psi$ is the boundary attraction term. The result of segmentation is shown in Figure 7(c).

Finally, candidates that are too small or too big to be mitotic nuclei were filtered out based on a minimum area of 200 pixels (that is $37 \mu\text{m}^2$, corresponding roughly to a circle of diameter $6.86 \mu\text{m}$) and a maximum area of 5,500 pixels (that is $1,017.5 \mu\text{m}^2$, corresponding roughly to a circle of diameter $36 \mu\text{m}$), computed on the segmented regions from MITOS multispectral dataset. An example of candidate detection is shown in Figure 7(d).

The process of cell division has four different stages, each one exhibiting different shape, size and textures. This motivates further statistical and morphological analysis on multispectral data to achieve reasonable classification of candidates into mitotic and non-mitotic types.

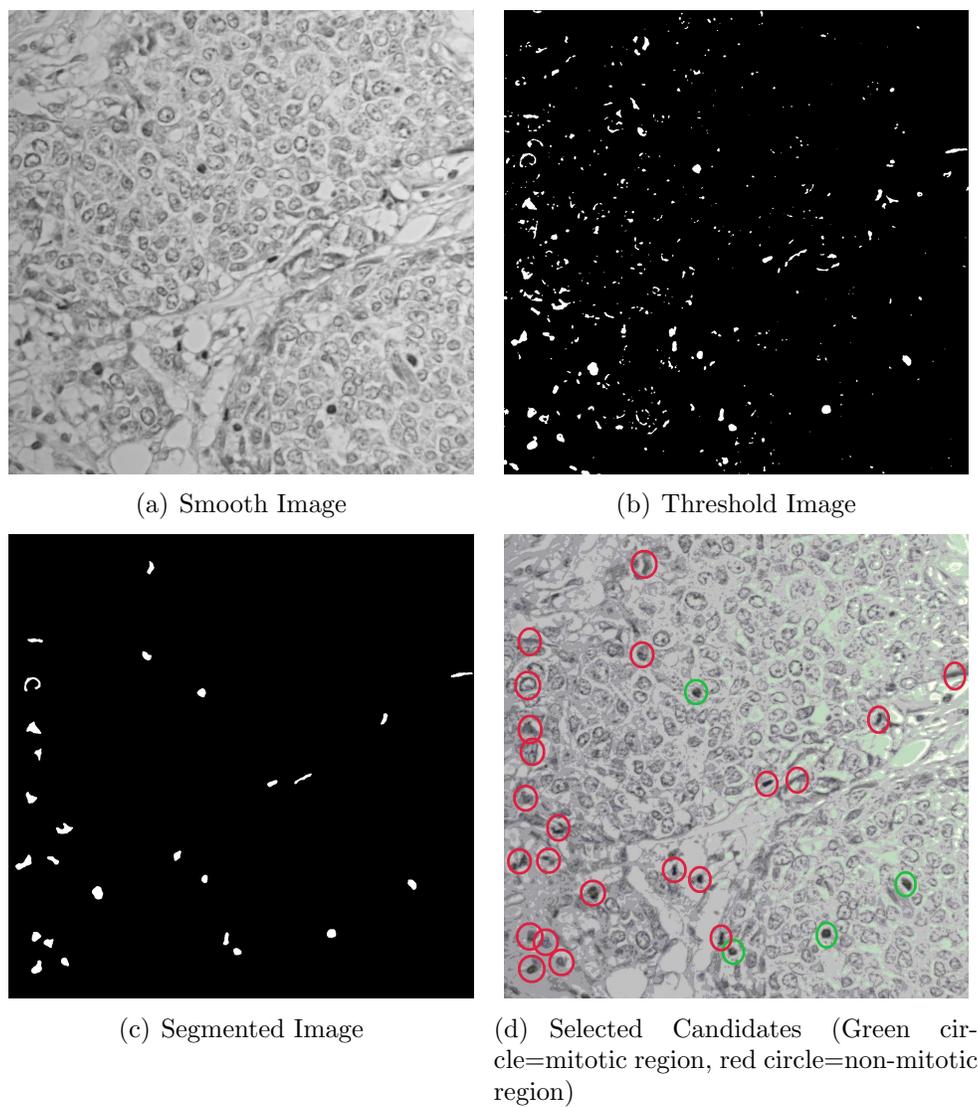


Figure 7: Different steps in candidate detection on breast cancer multispectral histopathology.

4.3. Morphological & Multispectral Statistical Features (MMSF) Computation

We computed MMSF vector consisting of intensity and textural features in several selected spectral bands. In addition, we also computed morphological features (such as area, roundness, elongation, perimeter and equivalent

spherical perimeter) from segmented regions during candidate detection. The morphological features reflect the phenotype information of mitotic and non-mitotic nuclei.

Using pixel information in selected spectral bands, we computed five intensity-based features on each segmented candidate regions in selected spectral bands: mean, median, variance, kurtosis and skewness for each candidate. For capturing spatial dependency inside candidate regions, we also computed multispectral textural features on each segmented candidate regions in selected spectral bands that included Haralick co-occurrence [52] and run-length [53] features. The co-occurrence features were correlation, cluster shade, cluster prominence, energy, entropy, hara-correlation, inertia and difference moment and the run-length features were short run emphasis, long run emphasis, grey-level non-uniformity, run length non-uniformity, low grey level runs emphasis, high grey level runs emphasis, short run low grey level emphasis, short run high grey level emphasis, long run low grey level emphasis, and long run high grey level emphasis. These multispectral texture features were computed with one displacement vector in four directions ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) as in [11]. To make these features rotationally invariant, they were averaged over four directions. Hence, a total of 23 features were computed for each spectral band (five intensity based features, eight co-occurrence features and ten run-length features).

In most cases, the features have different dynamic ranges. These different dynamic ranges of features affect the majority of classifiers which use the distance between two points. If one of the features has a broad range of values, the distance will be governed by this particular feature. Therefore, dynamic ranges of features were normalized so that each feature contributes approximately proportionately to the final distance. We solved this problem by normalizing the features values so that they lie within similar dynamic ranges. The normalization formula is given as:

$$f' = \frac{f - f_{\min}}{f_{\max} - f_{\min}} \quad (9)$$

where f is original feature value, f' is the normalized feature value, f_{\min} is the minimum feature values and f_{\max} is the maximum feature value.

4.4. Candidate Classification

Different classifiers learn a model in different ways that fits the relationship between features and labels of the training set. These learned models

are used to classify new unknown instances from the evaluation set. Each classifier has its own strong and weak points. To explore the different learning models on MMSF, four classifiers, DT, MLP, L-SVM and NL-SVM [54] were used to learn mitotic and non-mitotic classes from training set. These learned classes predicted the possible label of detected candidates in evaluation set. Generally, SVM and MLP tend to perform better when dealing with high dimensions and continuous features. On the other hand, DT tend to perform better when dealing with discrete features.

In our experiments, a multilayer perceptron (feedforward neural network) with backpropagation as learning model and sigmoid as activation function is used. The decision tree (DT) is a functional tree that could have logistic regression model at the inner nodes and leaves. It models the posterior class probabilities as:

$$P(C = c, X = x) = \frac{e^{F_c(x)}}{\sum_{i=1}^C e^{F_i(x)}} \quad (10)$$

where C is a label set and X is an instance set.

The L-SVM is a L2-regularized L2-loss support vector classification [55] as:

$$\min_{\omega} \quad \frac{1}{2}\omega^T\omega + \alpha \sum_{i=1}^N (\max(0, 1 - C_i\omega^T x_i))^2 \quad (11)$$

where $\alpha > 0$ is a penalty parameter, x_i is an instance of the instance set, C_i is a label of label set and ω is a weight vector.

In NL-SVM, we used Gaussian radial base function that allowed to avoid the explicit evaluation of mapping by using the so-called kernel trick:

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}} \quad (12)$$

5. Experiment Results

The proposed framework was evaluated on MITOS multispectral dataset [50]. The results of candidate detection and classification methods were compared with ground-truth (GT) information provided along with the dataset. The metrics used to evaluate the mitosis detection included: number of true positives (TP), number of false positives (FP), number of false negatives (FN), sensitivity or true positive rate ($\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$), precision or positive predictive value ($\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$) and F-Measure ($\text{F-Measure} = 2 \times \frac{\text{TPR} \times \text{PPV}}{\text{TPR} + \text{PPV}}$).

In addition to MITOS contest, the proposed framework was also evaluated with 5-fold cross validation [56] by merging the training and evaluation sets.

5.1. Spectral Bands Selection

How many spectral bands are necessary for a good detection of mitotic figures? Which spectral bands are relevant for mitotic figure detection?

To discuss these two questions, we tried first to evaluate the contributions of each spectral band using three different proposed methods as discussed in section 4.1. The results are shown in Table 2. The ranking of spectral bands in method one is based on difference between spectral absorption of mitotic nuclei and cytoplasm, while the ranking of spectral bands in method two is based on difference between hematoxylin and eosin spectral absorption. These three rankings put the same spectral bands, namely 7, 8 and 9 in top position. More specifically, the top position in methods one and two is occupied by spectral band 7 while method three gives the top position to spectral band 8 on the basis of highest MI.

At the bottom of the table, there are spectral bands 4 and 5 for all three rankings. According to method two ranking, the difference between absorption response of hematoxylin and eosin in spectral bands 4 and 5 is almost zero which represents that these two spectral bands are irrelevant for mitosis discrimination. Based on these analyses, we ignore spectral bands 4 and 5 for mitosis discrimination.

Considering the available spectral bands and their rankings, our selection of spectral bands contains the following eight bands: 8, 9, 7, 6, 2, 1, 3, and 0.

5.2. Candidate Detection

We performed candidate detection in top three spectral bands of the three proposed spectral bands selection methods. Those three spectral bands are 7, 8 and 9. In order to evaluate the ability of these spectral bands to provide adequate information for the detection of mitosis, we will also perform separately candidate detection on spectral band 1 only as this band covers the whole visible spectrum. The results of candidate detection step were ranked according to F-Measure and reported in Figure 8.

Spectral band 8 has higher F-Measure than bands 7, 9 and 1 with more TP and less FP. Although spectral band 1 covers the full spectrum of light, it reported poor results for candidate detection.

On training and evaluation sets, the candidate detection using spectral band 8 detects 2510 and 1188 candidates, containing 192 and 90 ground truth

Table 2: Different Rankings of Spectral Bands (SBs). The upper dotted line shows that SBs 7,8 and 9 are at top three positions in these ranking. The lower dotted line shows that SBs 4 and 5 are at bottom three positions.

Method 1		Method 2		Method 3	
SB	Mitosis–Cytoplasm	SB	H–E	SB	MI
7	0.47	7	0.96	8	3.6
8	0.45	8	0.91	9	3.59
9	0.36	9	0.64	7	3.38
3	0.33	1	0.39	6	3.18
2	0.31	6	0.33	2	3.16
6	0.30	0	0.23	1	3.05
1	0.30	2	0.23	3	3.05
4	0.29	3	0.21	0	2.99
0	0.28	5	0.04	4	2.95
5	0.27	4	0	5	2.83

mitosis from a total 224 and 98 ground truth mitosis, respectively. Among all the detected candidates, there are 2318 and 1098 non-mitosis in the training and evaluation sets, respectively. The candidate detection step generates a large number of FP and missed 32 and 8 ground truth mitosis from training and evaluation sets, respectively.

5.3. Candidate Classification

5.3.1. Classification on MITOS Dataset using 5-fold Cross Validation

In this experiment, the assessment of classification performance was made using 5-fold cross validation by combining both MITOS training and evaluation sets. Classification results are shown in Table 3. In case of all spectral bands MMSF, NL-SVM classifier outperformed the other classifiers and achieved the highest PPV (79.70%) and F-Measure (61.45%). Overall, L-SVM reported higher TPR (56.21%) but more FP as well. In case of selected eight spectral bands MMSF, L-SVM classifier has its F-Measure improved thanks to higher TPR and PPV. Overall, NL-SVM achieved the highest F-Measure (62.07%) and PPV (81.00%).

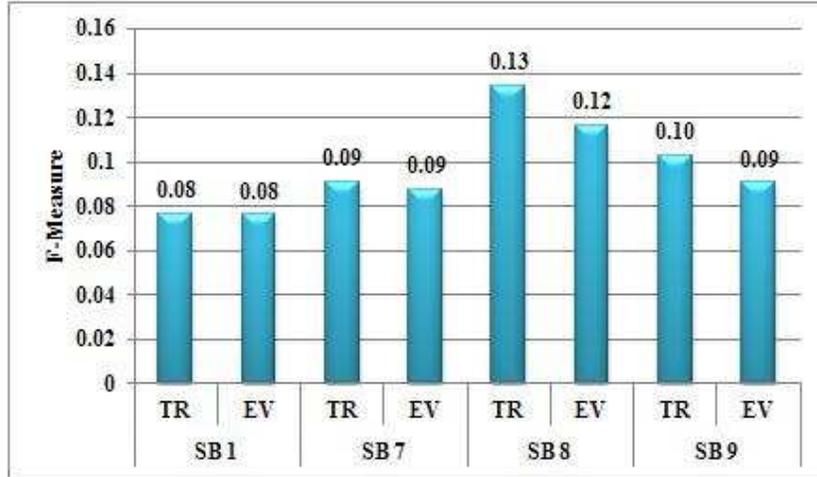


Figure 8: Candidate detection results on selected spectral bands (SB: spectral band, TR: training dataset, EV: evaluation dataset).

Table 3: Classification Results (ground truth = 322 mitosis) on MITOS Dataset using 5-fold Cross Validation.

Classifiers	TP	FP	TPR	PPV	F-Measure
All SBs MMSF					
L-SVM	181	98	56.21%	64.87%	60.23%
NL-SVM	161	41	50%	79.70%	61.45%
MLP	164	63	50.93%	72.25%	59.54%
DT	167	103	51.86%	56.42%	56.42%
Selected eight SBs MMSF					
L-SVM	184	93	57.14%	66.43%	61.44%
NL-SVM	162	38	50.31%	81.00%	62.07%
MLP	164	47	50.93%	77.73%	61.54%
DT	163	79	50.62%	67.36%	57.80%

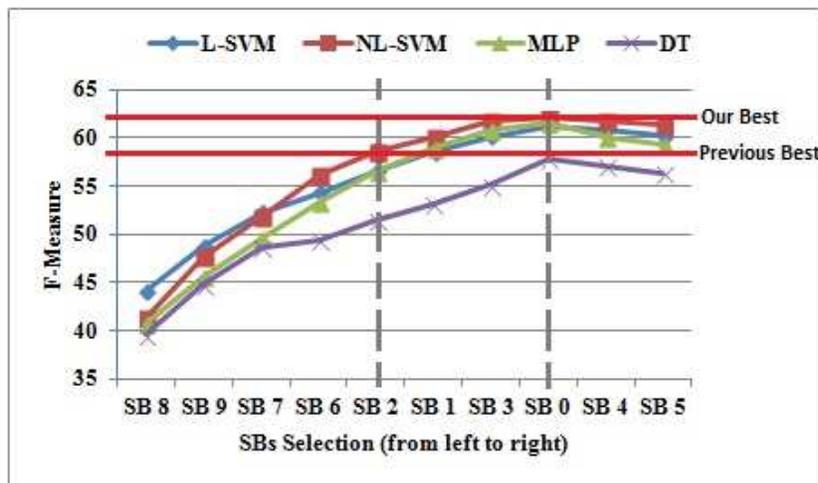


Figure 9: Plot of F-Measure using spectral bands (SBs) selection. Result from using all SBs from left to the current, e.g. SB 2 result uses SB 8, 9, 7, 6, 2. This order is taken from the mRMR ranking. First vertical dotted line shows that selecting features upto SBs 2 matches the previous best result. Second vertical dotted line highlights the overall best result by selecting features upto SBs 0.

Using mRMR ranking of spectral bands, the different selection of spectral bands are also tested using all the classifiers introduced in Section 4.4. However, for both clarity and brevity, we only plot F-Measure curves of 5-fold cross validation with all classifiers in Figure 9. Similar results were obtained in other cases. Figure 9 shows that F-Measure increases while we add more spectral bands to the set of selected spectral bands. F-Measure reaches a peak with a set of eight selected spectral bands, then it starts decreasing when adding more spectral bands. The sequence of spectral bands selection is according to ranking of MI from mRMR. In case of few spectral bands MMSF, NL-SVM, MLP and DT reported poor classification accuracy while L-SVM reported higher classification accuracy. As more spectral bands are selected NL-SVM classifiers started performing better than other classifiers and reached maximum performance with first eight selected spectral bands.

5.3.2. Classification on MITOS Evaluation Set

In this experiment, MITOS training set was used to train the four selected classifiers and MITOS evaluation set is used to test the classification accuracy of proposed framework as shown in Table 4. Classification was performed from MMSF computed both on selected eight spectral bands and on all the

Table 4: Classification Result (ground truth = 98 mitosis) on MITOS Evaluation Set.

Classifiers	TP	FP	TPR	PPV	F-Measure
All SBs MMSF					
L-SVM	62	41	63.27%	60.19%	61.69%
NL-SVM	53	25	54.08%	67.95%	60.23%
MLP	63	50	64.29%	55.75%	59.72%
DT	66	59	67.35%	52.80%	59.19%
Selected eight SBs MMSF					
L-SVM	63	38	64.29%	62.38%	63.32%
NL-SVM	58	26	59.18%	69.05%	63.74%
MLP	59	30	60.20%	66.29%	63.10%
DT	61	37	62.24%	62.24%	62.24%

spectral bands. When all spectral bands MMSF were used for training and evaluation, DT classifier reported higher TP (66) but also higher FP (87), resulting in low F-Measure (55.59%). The NL-SVM classifiers have few FP but also few TP resulting also in low F-Measure (57.71%). Overall, the MLP classifier reported the highest F-Measure (59.72%).

In case of selected spectral bands MMSF, F-Measure was increased more in NL-SVM classifiers by 42.22% reduction of FP and reported the highest F-Measure (63.74%). Overall, L-SVM classifier achieved the highest TPR (64.29%) while MLP classifier reported the second highest F-Measure (63.10%). The selection of spectral bands improved the classification accuracy by reducing the number of FP. This shows that the spectral bands not selected are not relevant for mitosis classification and therefore they may bias the classifiers towards non-mitosis class.

To consider potential over-fitting of classification, we have also tested our proposed framework using a training set made up from three out of the five slides and an evaluation set made up of the remaining two slides. LSVM classifier outperformed DT, MLP and NLSVM classifiers with TPR (80%), PPV (55%) and F-Measure (65.20%) on all spectral bands and TPR (74%), PPV (63%) and F-Measure (68.10%) on selected eight spectral bands. These results eliminate the potential risk of over-fitting of classification. In addition, by reducing the number of features, the classification results are of lower quality. This demonstrates that there is no over-fitting in our framework when using the complete set of features from selected spectral bands.

Table 5: Classification Result on MITOS Dataset with spectral band 1 vs other spectral bands (ground truth = 322 mitosis) using 5-fold cross validation.

Classifiers	TP	FP	TPR	PPV	F-Measure
White SB MMSF					
L-SVM	76	62	23.60%	55.07%	33.04%
NL-SVM	68	37	21.12%	65.38%	31.92%
MLP	72	41	22.36%	63.72%	33.10%
DT	59	30	18.32%	66.29%	28.71%
All SBs except band 1 MMSF					
L-SVM	179	111	55.59%	61.72%	58.50%
NL-SVM	160	48	49.69%	76.92%	60.38%
MLP	162	64	50.31%	71.68%	59.12%
DT	168	97	52.17%	63.40%	57.24%

5.3.3. Classification on MITOS Dataset with Spectral Band 1 vs other Spectral Bands

To investigate the relative contribution of spectral bands, we also performed a comparative study of mitosis classification using MMSF features from spectral band 1 (white spectral band) with MMSF features from rest of spectral bands and the achieved results are shown in Table 5. It is important to note that for all classifiers detection results with MMSF excluding spectral band 1 outperform detection results of MMSF on spectral band 1 only, and reports highest F-Measure (60.38%) with NL-SVM classifier. This experiment illustrates that multispectral bands have much more information for mitosis classification than white band alone.

5.3.4. Classification on MITOS Dataset with Blue, Green and Red Spectral Bands

In this experiment, we explored the impact of different parts of visible spectrum on the mitosis classification. The range of visible spectrum was divided into three parts: blue spectrum (spectral bands 2,3,4), green spectrum (spectral bands 5,6,7) and red spectrum (spectral bands 8,9,0). Each spectrum MMSF were used for mitosis classification and results are shown in Table 6. Green MMSF reported the highest TPR (46.89%) with L-SVM classifier and the highest PPV (75.14%) with NL-SVM classifier, while red MMSF reported the highest F-Measure (54.58%) with L-SVM classifier. These re-

Table 6: Classification Result on MITOS Dataset with blue, green and red Spectral Bands (SBs) using 5-fold cross validation (ground truth = 322 mitosis).

Classifiers	TPR	PPV	F-Measure	TPR	PPV	F-Measure
	Blue SBs (2,3&4) MMSF			Green SBs (5,6&7) MMSF		
L-SVM	26%	59%	36.56%	47%	56%	51.01%
NL-SVM	32%	68%	43.31%	40%	67%	50.49%
MLP	39%	55%	45.26%	43%	60%	49.91%
DT	30%	51%	37.96%	42%	55%	47.89%
	Red SBs (0,8&9) MMSF			SBs 7,8&9 MMSF		
L-SVM	42%	67%	51.53%	45%	62%	52.25%
NL-SVM	39%	72%	51%	41%	72%	51.88%
MLP	42%	61%	49.54%	40%	66%	49.62%
DT	40%	58%	47.43%	41%	61%	48.70%

sults illustrate that red spectral bands have more information for mitosis classification as compared to green and blue spectral bands. An important point is that green spectral bands have more information as compared to blue ones that might be due to spectral bands 6 and 7 having higher absorption response for mitotic figures. Moreover, we also investigated the impact of top three ranked spectral bands (7, 8 & 9) MMSF vector for mitosis detection using 5-fold cross validation. As compared with blue, green and red spectral bands results, this MMSF vector clearly outperformed with all four classifiers and achieved highest F-Measure (52.25%) with L-SVM classifier as shown in Table 6.

6. Discussion

The results seem to indicate the best scores are achieved using the selected focal plane, across selected spectral bands. The proposed focal plane selection is automatic and unsupervised. We performed selection of focal plane separately for each spectral band. However, as the best focal plane is the same for all spectral bands, in future we could use the best focal plane computed on one spectral band only. It is not necessary to compute it for each spectral band separately. In other words, finding the best focal plane and finding the best spectral bands are separable problems.

The best F-Measure for candidate detection was achieved on spectral band 8. The fact that the proposed framework achieved better results when

using spectral bands 8, 9 and 7 for candidate detection than when using the full spectrum (spectral band 1) supports the claim that multispectral imaging improves the accuracy of the framework. As spectral bands 7, 8 and 9 actually overlap in terms of spectrum, it would be interesting to apply spectral unmixing between spectral bands 7, 8 and 9 to see if it can further improve the results. The results illustrate clearly the improved accuracy resulting of the spectral band selection process.

In order to study the MMSF for classification of mitotic figures, we performed spectral bands selection by studying which spectral bands have minimum redundant and maximum relevant information for mitosis classification. Figure 9 show the results (F-Measure) of mitosis classification using a selection of spectral bands based on mRMR ranking. With top five spectral bands, we match the best results achieved during MITOS contest [46]. With top eight spectral bands, we achieve 63.74% F-Measure which is a 8% increase over the best know results. Adding more spectral bands (spectral bands 4 and 5), only has a negative impact on the results. This concludes that spectral bands 4 and 5 are irrelevant for mitosis discrimination. To illustrate the possible correlation between selected spectral bands and the staining characteristics of the spectral samples, the shape of H–E plot in Figure 6 validated the proposed selection of spectral bands.

We also used the analysis of different subsets of multispectral features as complement to the analysis of performance on MMSF. Specifically, we have shown in experiment 3 that multispectral data contains more discriminant information for detection of mitosis than spectral band 1, which covers the full visible spectrum. According to experiment 4, red spectral bands (8,9,0) are more helpful for mitosis discrimination as compared to blue spectral bands (2,3,4) and green spectral bands (5,6,7). An important finding is that green spectral bands have more information than blue spectral bands, which might be reason of spectral band 6 and 7, which contain higher absorption response for mitotic nuclei as compared to other tissue components.

Separate training and evaluation sets of MITOS dataset [50] have been used for training and evaluation of the proposed framework. The comparison of proposed framework results with MITOS contest results [46] are shown in Figure 10. Malon and Cosatto [34] method ranked first of the contest with 59% F-Measure. In comparison with MITOS contestants, the proposed framework compute features on selected spectral bands that have higher mitotic absorption response as compared to other tissue components. Using selected spectral bands MMSF and NL-SVM classifier, we managed to achieve

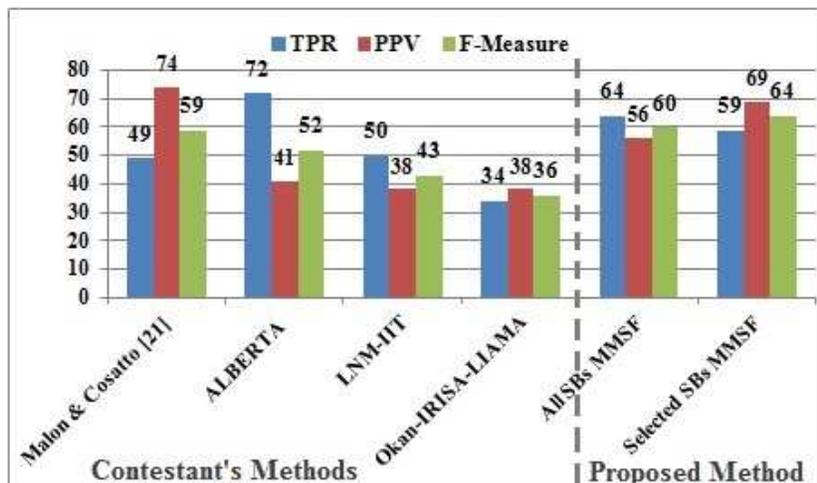


Figure 10: Comparison of proposed framework results with MITOS contest result. Vertical dotted line is used to separate the result of contestants methods and proposed method.

highest F-Measure (63.74%). Fig. 11 illustrates the receiver operating characteristic (ROC) curve [57] obtained with selected spectral bands MMSF with NL-SVM classifier. This clearly demonstrates that our new proposed framework results in an improved ability to distinguish mitosis from other objects. The proposed framework should be considered as the state of the art.

The majority of mitosis missed by the proposed framework, are significantly different from the detected mitosis. Missed mitosis have very light nuclear material and a small size as compared to detected mitosis. In addition, some missed mitotic figures are clustered in small heterogeneous regions. Furthermore, most of FP are lymphocytes that look very similar to mitotic figures.

The proposed framework is developed using Insight Segmentation and Registration Toolkit (ITK) [58] and it requires an average of 10 minutes to process an image of 1360×1360 pixels on Intel(R) Core(TM) i5-2500 CPU @ 3.30GHz with 4GB memory.

7. Conclusion

An automated mitosis detection framework for breast cancer multispectral imaging based on multispectral spatial features has been proposed. First,

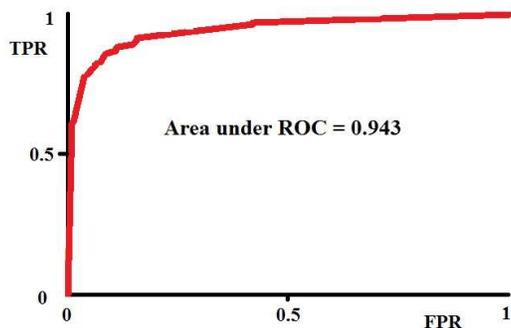


Figure 11: The ROC curve using selected spectral bands MMSF with NL-SVM classifier.

focal plane selection is performed using maximum gradient information. Based on MI of spectral bands and spectral absorption of different tissue components and stains, spectral bands were selected for candidate detection and feature computation. Candidate detection was performed on the spectral bands that have relatively higher MI and mitotic absorption spectra. Then, MMSF are computed for each candidate in eight selected spectral bands, a highly efficient model for capturing spectral and spatial features for object discrimination. The proposed framework outperformed the MITOS contest results with 8% improvement of F-Measure. We expect to improve mitosis detection performance by selecting the feature set through the computation of MMSF of candidate regions in selected spectral bands. In future work, we plan to investigate unmixing of bands as most spectral bands have overlapping area, which increase redundancy. The pre-selection of the focal plane (or volumes) is also of great importance to reduce the complexity of the dataset and improve the actual performances to reach clinical operational acceptance expected by our professional consortia. Future improvements are expected using this novel medical modality in histopathology, in order to exploit the richness of the spectral and spatial information, going far beyond the human eye perception.

Acknowledgement

This work is supported by the French National Research Agency ANR, project MICO, reference ANR-10-TECS-015.

References

- [1] J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. Parkin, D. Forman, F. Bray, GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet], International Agency for Research on Cancer, Lyon, France, 2013, available from: <http://globocan.iarc.fr/>, accessed on December 20th, 2013
- [2] H. Bloom, W. Richardson, Histological grading and prognosis in breast cancer: a study of 1409 cases of which 359 have been followed for 15 years, *British Journal of Cancer* 11 (3) (1957) 359.
- [3] E. Elston, I. Ellis, Method for grading breast cancer., *Journal of clinical pathology* 46 (2) (1993) 189.
- [4] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, A. Tozeren, Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer, *BMC Medical Imaging* 6 (1) (2006) 14.
- [5] 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: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, IEEE, 2008, pp. 284–287.
- [6] E. Cosatto, M. Miller, H. P. Graf, J. S. Meyer, Grading nuclear pleomorphism on histological micrographs, in: *19th International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1–4.
- [7] J.-R. Dalle, H. Li, C.-H. Huang, W. K. Leow, D. Racoceanu, T. C. Putti, Nuclear pleomorphism scoring by selective cell nuclei detection, in: *IEEE Workshop on Applications of Computer Vision*, 2009.
- [8] A. R. Chaudhury, K. Iychettira, R. Iyer, A. Sreedevi, Diagnosis of invasive ductal carcinoma using image processing techniques, in: *International Conference on International Conference on Image Information Processing (ICIIP)*, IEEE, 2011, pp. 1–6.
- [9] M. Dunder, S. Badve, G. Bilgin, V. C. Raykar, R. K. Jain, O. Sertel, M. N. Gurcan, Computerized classification of intraductal breast lesions

- using histopathological images, *IEEE Transactions on Biomedical Engineering* 58 (7) (2011) 1977–1984.
- [10] H. Irshad, S. Jalali, L. Roux, D. Racoceanu, L. Joo-Hwee, G. L. Naour, F. Capron, Automated mitosis detection using texture, sift features and hmax biologically inspired approach, *Journal of Pathology informatics* 4 (2013) 12. doi:10.4103/2153-3559.109870.
 - [11] H. Irshad, Automated mitosis detection in histopathology using morphological and multi-channel statistics features, *Journal of Pathology Informatics* 4 (2013) 10. doi:10.4103/2153-3539.112695.
 - [12] H. Irshad, A. Veillard, L. Roux, D. Racoceanu, Methods for Nuclei Detection, Segmentation and Classification in Digital Histopathology: A Review. Current Status and Future Potential, *IEEE Reviews on Biomedical Engineering (RBME)*, 2013, issue 99, pp. 1.
 - [13] G. Avwioro, Histochemical uses of haematoxylina review, *J Pharm Clin Sci (JPCS)* 1 (2011) 24–34.
 - [14] G. Meijer, J. Belien, P. Van Diest, J. Baak, Origins of image analysis in clinical pathology, *Journal of clinical pathology* 50 (5) (1997) 365–370.
 - [15] T. J. Fuchs, J. M. Buhmann, Computational pathology: Challenges and promises for tissue analysis, *Computerized Medical Imaging and Graphics* 35 (7) (2011) 515–530.
 - [16] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, B. Yener, Histopathological image analysis: A review, *IEEE Reviews in Biomedical Engineering* 2 (2009) 147–171.
 - [17] W. H. Wolberg, W. N. Street, O. L. Mangasarian, Breast cytology diagnosis via digital image analysis, *Analytical and Quantitative Cytology and Histology* 15 (6) (1993) 396–404.
 - [18] R. N. Clark, T. L. Roush, Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications, *Journal of Geophysical Research: Solid Earth* (19782012) 89 (B7), 1984, pp. 6329-6340.
 - [19] J. R. Jensen, D. C. Cowen, Remote sensing of urban/suburban infrastructure and socio-economic attributes, *Photogrammetric engineering and remote sensing*, vol. 65, 1999, pp. 611-622.

- [20] V. Walter, Object-based classification of remote sensing data for change detection, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58 (3), 2004, pp. 225-238.
- [21] T. M. Lillesand, R. W. Kiefer, J. W. Chipman, *Remote sensing and image interpretation*, ed 5, John Wiley & Sons Ltd, 2004.
- [22] D. Zhang, Z. Guo, G. Lu, L. Zhang, W. Zuo, An online system of multispectral palmprint verification, *Instrumentation and Measurement, IEEE Transactions on* 59 (2), 2010, pp. 480-490.
- [23] D. Wei, L. Zhang, D. Zhang, Q. Pan, Studies on hyperspectral face recognition in visible spectrum with feature band selection, *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 40 (6), 2010, pp. 1354-1361.
- [24] P. S. Thenkabail, R. B. Smith, D. E. Pauw, Hyperspectral vegetation indices and their relationships with agricultural crop characteristics, *Remote sensing of Environment*, vol. 71 (2), 2000, pp. 158-182.
- [25] Y. Li, Y. Zhu, Y. Tian, W. Cao, Quantitative relationships between leaf area index and canopy reflectance spectra of wheat, *The journal of applied ecology*, vol. 17 (8), 2006, 1443.
- [26] S. Prigent, X. Descombes, D. Zugaj, L. Petit, A. S. Dugaret, P. Martel, J. Zerubia, Skin lesion evaluation from multispectral images, *Research report*, 2012.
- [27] J. Feng, H. Li, J. J. Duan, X. Lv, Y. Zhu, W. Liu, Multispectral Bands Selection of Horticultural Plant Diseases Using Brightness, *Photonics and Optoelectronics (SOPO), 2012 Symposium on*, pp. 1-4. IEEE, 2012.
- [28] J. Holmquist, Y. Imasoto, E. Bengtsson, B. Olsen, B. Stenkvist, A microspectrophotometric study of Papanicolaou-stained cervical cells as an aid in computerized image processing, *Journal of Histochemistry & Cytochemistry*, 24 (12), 1976, pp. 1218-1224.
- [29] P. Geladi, H. Grahn, K. Esbensen, E. Bengtsson, Image analysis in chemistry II. Multivariate image analysis. *TRAC Trends In Analytical Chemistry*, 11(3), 1992, pp. 121-130.

- [30] C. Stewart, J. Duncan, M. Farquharson, J. Richmond, Fine needle aspiration cytology diagnosis of malignant lymphoma and reactive lymphoid hyperplasia., *Journal of clinical pathology* 51 (3) (1998) 197–203.
- [31] E. S. Cibas, B. S. Ducatman, *Cytology: Diagnostic principles and clinical correlates*, Saunders, 2009.
- [32] M. E. Plissiti, C. Nikou, A review of automated techniques for cervical cell image analysis and classification, in: *Biomedical Imaging and Computational Modeling in Biomechanics*, Springer, 2013, pp. 1–18.
- [33] Y. Gong, *Breast cancer: Pathology, cytology, and core needle biopsy methods for diagnosis*, in: M. K. Shetty (Ed.), *Breast and Gynecological Cancers*, Springer New York, 2013, pp. 19–37.
- [34] C. Malon, E. Cosatto, et al., Classification of mitotic figures with convolutional neural networks and seeded blob features, *Journal of Pathology Informatics* 4 (2013) 9. doi:10.4103/2153-3539.112694.
- [35] R. M. Levenson, Spectral imaging perspective on cytomics, *Cytometry Part A* 69 (7) (2006) 592–600.
- [36] R. M. Levenson, A. Fornari, M. Loda, Multispectral imaging and pathology: seeing and doing more, *Expert Opinion on Medical Diagnostics* 2 (9) (2008) 1067–1081, pMID: 23495926.
- [37] D. C. Fernandez, R. Bhargava, S. M. Hewitt, I. W. Levin, Infrared spectroscopic imaging for histopathologic recognition, *Nature biotechnology* 23 (4) (2005) 469–474.
- [38] R. M. Levenson, J. R. Mansfield, Multispectral imaging in biology and medicine: slices of life, *Cytometry Part A* 69 (8) (2006) 748–758.
- [39] X. Wu, M. Amrikachi, S. K. Shah, Embedding topic discovery in conditional random fields model for segmenting nuclei using multispectral data, *IEEE Transactions on Biomedical Engineering* 59 (6) (2012) 1539–1549.
- [40] R. Khelifi, M. Adel, S. Bourennane, Multispectral texture characterization: application to computer aided diagnosis on prostatic tissue images, *EURASIP Journal on Advances in Signal Processing* 2012 (1) (2012) 1–13.

- [41] R. M. Levenson, P. J. Cronin, K. K. Pankratov, Spectral imaging for brightfield microscopy, in: *Biomedical Optics 2003*, International Society for Optics and Photonics, 2003, pp. 27–33.
- [42] S. M. Gentry, R. M. Levenson, Biomedical applications of the information-efficient spectral imaging sensor (ISIS), in: *International Biomedical Optics Symposium (BiOS)*, International Society for Optics and Photonics, 1999, pp. 129–142.
- [43] M. Roula, A. Bouridane, F. Kurugollu, A. Amira, A quadratic classifier based on multispectral texture features for prostate cancer diagnosis, in: *Seventh International Symposium on Signal Processing and Its Applications*, Vol. 2, IEEE, 2003, pp. 37–40.
- [44] K. Masood, N. Rajpoot, Texture based classification of hyperspectral colon biopsy samples using clbp, in: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro ISBI*, IEEE, 2009, pp. 1011–1014.
- [45] L. E. Boucheron, Z. Bi, N. R. Harvey, B. Manjunath, D. L. Rimm, Utility of multispectral imaging for nuclear classification of routine clinical histopathology imagery, *BMC Cell Biology* 8 (Suppl 1) (2007) S8.
- [46] L. Roux, D. Racoceanu, N. Lomenie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. L. Naour, M. N. Gurcan, Mitosis detection in breast cancer histological images an icpr 2012 contest, *Journal of Pathology Informatics* 4 (2013) 8. doi:10.4103/2153-3539.112693.
- [47] A. Martinez-Uso, F. Pla, J. Sotoca, P. Garcia-Sevilla, Clustering-based multispectral band selection using mutual information, in: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, Vol. 2, 2006, pp. 760–763. doi:10.1109/ICPR.2006.375.
- [48] M. Kamandar, H. Ghassemian, Maximum relevance, minimum redundancy band selection for hyperspectral images, in: *Electrical Engineering (ICEE), 2011 19th Iranian Conference on*, IEEE, 2011, pp. 1–5.
- [49] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy,, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 27 (8) (2005) 1226–1238.

- [50] MITOS contest 2012, <http://ipal.cnrs.fr/ICPR2012>.
- [51] P. A. Bautista, Y. Yagi, Digital simulation of staining in histopathology multispectral images: enhancement and linear transformation of spectral transmittance, *Journal of Biomedical Optics* 17 (5) (2012) 0560131–05601310.
- [52] R. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, *IEEE Trans. on Systems, Man and Cybernetics* 3 (1973) 610–21.
- [53] M. Galloway, Texture analysis using gray level run lengths, *CGIP* 4 (1975) 172–9.
- [54] Weka, Url <http://www.cs.waikato.ac.nz/~ml/weka>.
- [55] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, C.-J. Lin, Liblinear: A library for large linear classification, *The Journal of Machine Learning Research* 9 (2008) 1871–1874.
- [56] P. A. Devijver, J. Kittler, *Pattern recognition: A statistical approach*, Prentice/Hall International Englewood Cliffs, NJ, 1982.
- [57] M. H. Zweig, G. Campbell, Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine., *Clinical chemistry* 39 (4) (1993) 561–577.
- [58] N. C. Ibáñez, *The ITK Software Guide, First Edition*., published by Kitware Inc., 2003.
- [59] K. Mosaliganti, L. Cooper, R. Sharp, R. Machiraju, G. Leone, K. Huang, J. Saltz, Reconstruction of cellular biological structures from optical microscopy data, *IEEE Transactions on Visualization and Computer Graphics*, 2008, vol 14 (4), 863–876.

Appendix

A - Focal plane Selection using Maximum Gradient

For selection of focal plane, average gradient of mitotic figures from background regions was computed on all the focal planes. The computed gradient

vector of image I :

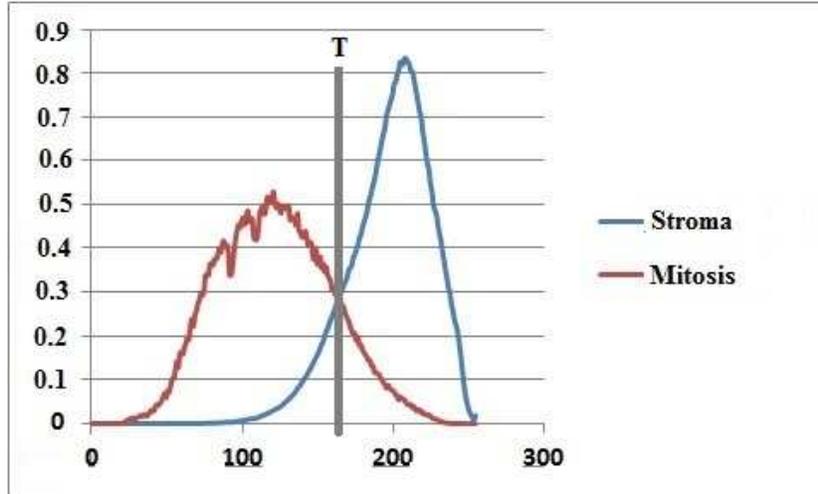
$$\nabla I = \left[\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right] \quad (13)$$

where $\frac{\partial I}{\partial x}$ and $\frac{\partial I}{\partial y}$ are partial derivative of I with respect to x and y directions, respectively as:

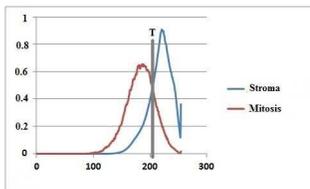
$$\begin{aligned} \frac{\partial I(x, y)}{\partial x} &= \frac{I(x+1, y) - I(x-1, y)}{2} \\ \frac{\partial I(x, y)}{\partial y} &= \frac{I(x, y+1) - I(x, y-1)}{2} \end{aligned} \quad (14)$$

The focal plane that has maximum average gradient (i.e. best focus) is selected for the next steps of the framework.

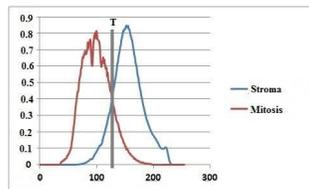
B - Histograms of mitotic and stroma regions in 10 spectral bands



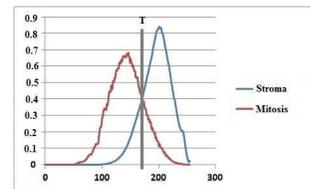
(a) Histogram in SB 8



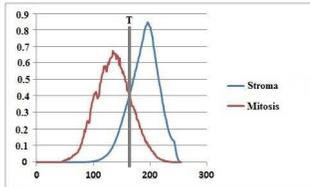
(b) Histogram in SB 0



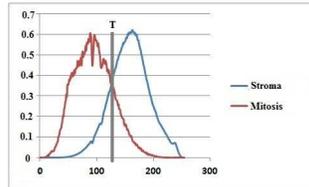
(c) Histogram in SB 1



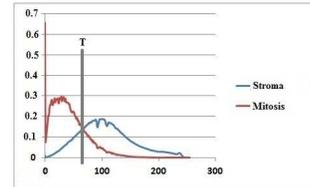
(d) Histogram in SB 2



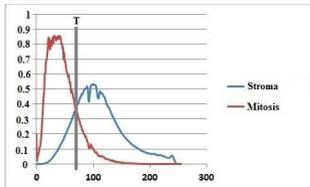
(e) Histogram in SB 3



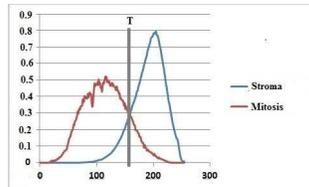
(f) Histogram in SB 4



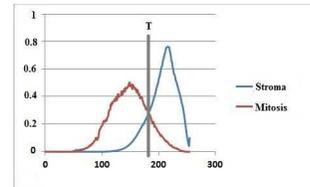
(g) Histogram in SB 5



(h) Histogram in SB 6



(i) Histogram in SB 7



(j) Histogram in SB 9

Figure 12: Histogram analysis of mitotic and stroma regions in 10 spectral bands (SBs). The threshold value T is selected at intersection of both curves.