

**REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
DEPARTMENT OF COMPUTER ENGINEERING**



**MITOSIS DETECTION IN MULTISPECTRAL
HISTOPATHOLOGIC IMAGES**

14011901 – Sardor HAZRATOV

SENIOR PROJECT

Advisor
Assoc. Prof. Dr. Gökhan BİLGİN

January, 2019

ACKNOWLEDGEMENTS

First, I want to thank my thesis supervisor, Gökhan BİLGİN for his guidance and feedback, as well as for providing me with the resources to carry out this work. His advice has been invaluable in helping me gain a deeper understanding of the subject.

I also thank my parents for their never-ending support throughout my studies, without which none of this would have been possible.

Sardor HAZRATOV

TABLE OF CONTENTS

LIST OF ABBREVIATIONS	v
LIST OF FIGURES	vi
LIST OF TABLES	vii
ABSTRACT	viii
ÖZET	ix
1 Introduction	1
1.1 Literature Review	1
1.2 Objective of the Thesis	1
1.3 Hypothesis	1
2 Related works	3
3 Feasibility	4
3.1 Technical Feasibility	4
3.2 Legal and Economic Feasibility	5
3.3 Time Feasibility	5
4 System Analysis	6
5 System Design	8
5.1 Image Data	8
5.2 Implementation of Discriminative Image	9
5.3 Implementation of CNN model	10
5.4 Generel overview of proposed technique	11
6 Application	12
7 Experimental Results	14
8 Performance Analysis	15

9 Conclusion	16
References	17
Curriculum Vitae	18

LIST OF ABBREVIATIONS

HP	Histopathologic
HPI	Histopathologic images
R&E	Hematoxylin and Eosin
MC	Mitotic cells
HPF	High Power Fields
LDA	Linear discriminant analysis
CNN	Convolutional neural network
DNN	Deep neural network
NN	Neural network
DI	Discriminative Image

LIST OF FIGURES

Figure 1.1	Spectral bands of the multi-spectral microscope	2
Figure 1.2	Mitotic cell on 10 different spectral bands	2
Figure 3.1	Time feasibility: Gantt diagram	5
Figure 4.1	Locations of quarters on multispectral images	7
Figure 4.2	Annotated mitosis on multispectral space	7
Figure 5.1	CNN model architecture	10
Figure 5.2	Generel steps flowchart	11
Figure 6.1	Sample image with 10 spectral bands and generated DI image	12
Figure 6.2	Samples for C1 class with 10 bands (mitosis regions)	13
Figure 6.3	Samples for C2 class with 10 band (non-mitosis regions)	13
Figure 7.1	Mitosis prediction result-1	14
Figure 7.2	Mitosis prediction result-2	14
Figure 7.3	Mitosis prediction result-3	14
Figure 7.4	Mitosis prediction result-4	14
Figure 7.5	Mitosis prediction result-5	14
Figure 7.6	Mitosis prediction result-6	14

LIST OF TABLES

Table 8.1 Confusion table	15
-------------------------------------	----

ABSTRACT

MITOSIS DETECTION IN MULTISPECTRAL HISTOPATHOLOGIC IMAGES

Sardor HAZRATOV

Department of Computer Engineering
Senior Project

Advisor: Assoc. Prof. Dr. Gökhan BİLGİN

In this work, segmentation of cellular structures in the multispectral histopathological images and possibility of the discrimination within normal and mitotic cells has been investigated. Extracting the mitotic cell from the histopathological image is a very challenging task. In the first step, Discriminative Image was acquired using Linear Discriminant Analysis. Discriminative Image was used to get mitotic cell candidates to train on Deep Learning Networks. Discriminative Images were applied filter in order to retrieve dark areas since mitotic cells were commonly dark pixels. After that k-means algorithm with two clusters were applied to separate background and unify darker pixel values. Then by thresholding image and contour detection regions of interest has been derived. With the help of Convolutional Neural Networks classification problem handled to predict mitotic and non-mitotic regions. ICPR-2012 Dataset has been used in both training and predicting steps. F-measure evaluated as 0.760, recall as 0.723, and precision as 0.802.

Keywords: Mitosis, histopathology, cell, cancer.

ÖZET

MULTİSPEKTRAL HİSTOPATOLOJİK GÖRÜNTÜLERDE MİTOZ TESPİTİ

Sardor HAZRATOV

Bilgisayar Mühendisliği Bölümü

Bitirme Projesi

Danışman: Doç. Dr. Gökhan BİLGİN

Bu çalışmada, multispektral histopatolojik görüntülerde hücresel yapıların segmentasyonu, normal ve mitozlu hücrelerde ayırt edebilme olasılığı araştırılmıştır. Mitozlu hücreyi histopatolojik görüntüden çıkarmak çok zor bir iştir. İlk adımda, Lineer Diskriminant Analizi kullanılarak Ayrımcı Görüntüler elde edildi. Ayrımcı Görüntüler, mitotik hücre adayları elemek ve Derin Öğrenme Ağlarıyla eğitmek için kullanılmıştır. Mitotik hücreler genellikle koyu pikseller olduğundan, koyu alanları elde etmek için Ayrımcı Görüntülerde filtre uygulandı. Arka plan ve mitotik adayları ayırmak için 2 kümeli k-means algoritması uygulanmıştır. Konvolüsyonel Sinir Ağları'nın yardımıyla mitotik ve mitotik olmayan alanları sınıflandırma problemi ele alındı. ICPR-2012 veriseti hem eğitim hem de tahmin aşamalarında kullanılmıştır. Yapay Sinir Ağı eğitimi mitotik ve mitotik olmayan bölgelerin örnekleri ile gerçekleştirılmıştır. F-ölçüsü 0.760, geri çağırma oranı 0.723 ve hassasiyet oranı 0.802 olarak değerlendirildi.

Anahtar Kelimeler: Mitoz, histopatoloji, hücre, kanser.

1

Introduction

1.1 Literature Review

Breast cancer is the most common cancer among women and a major cause of death worldwide. The standard approach for breast cancer diagnosis relies on visual inspection of histopathological (HP) samples stained with Hematoxylin and Eosin (H&E). Traditional approach of detecting mitotic cells (MC) is manually analyzing H&E stained tissue using high-power microscopy. Mitosis detection is exhausting and complex process of diagnosing every cells. Automizing this process using novel image processing and pattern recognition techniques could reduce the effort, time and its costs making it more affordable.

1.2 Objective of the Thesis

The objective of this thesis is making mitosis detection system using modern image processing and machine learning techniques based on Multi-spectral ICPR-2012 dataset. The success of the system will be evaluated with Precision, F-Measure and Detection Rate.

1.3 Hypothesis

Mitotic count is an important parameter in breast cancer grading as it gives an evaluation of the aggressiveness of the tumour. Detection of mitosis is a very challenging task since in images, they appear as small objects which have a large variety of shapes. The four main phases of a mitosis are prophase, metaphase, anaphase and telophase. The shape of the nucleus is very different depending on the phase of the mitosis. On its last stage, the telophase, a mitosis has two distinct nuclei, but they are not yet full individual cells. A mitosis in telophase must be counted as one single mitosis, it should not be miscounted as two mitosis.[1]

Multi-spectral image dataset itself has 10 spectral band. The spectral bands are all in the visible spectrum. (see Figures 1.1, 1.2)

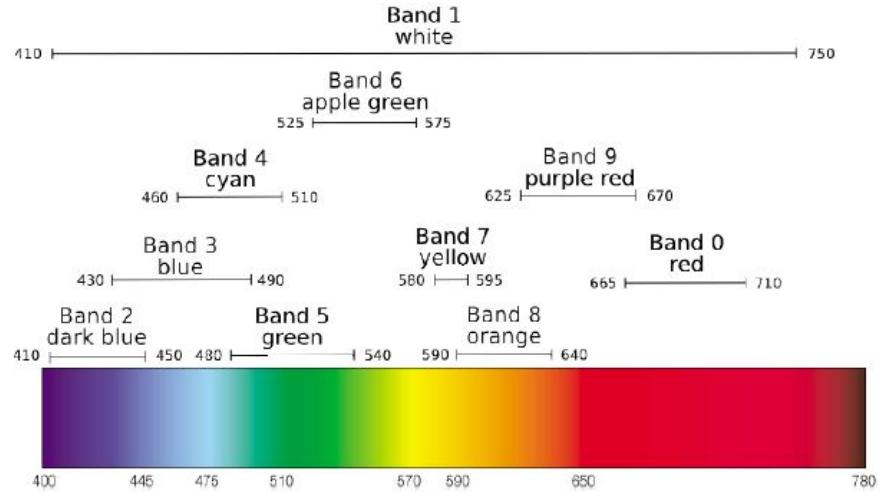


Figure 1.1 Spectral bands of the multi-spectral microscope.

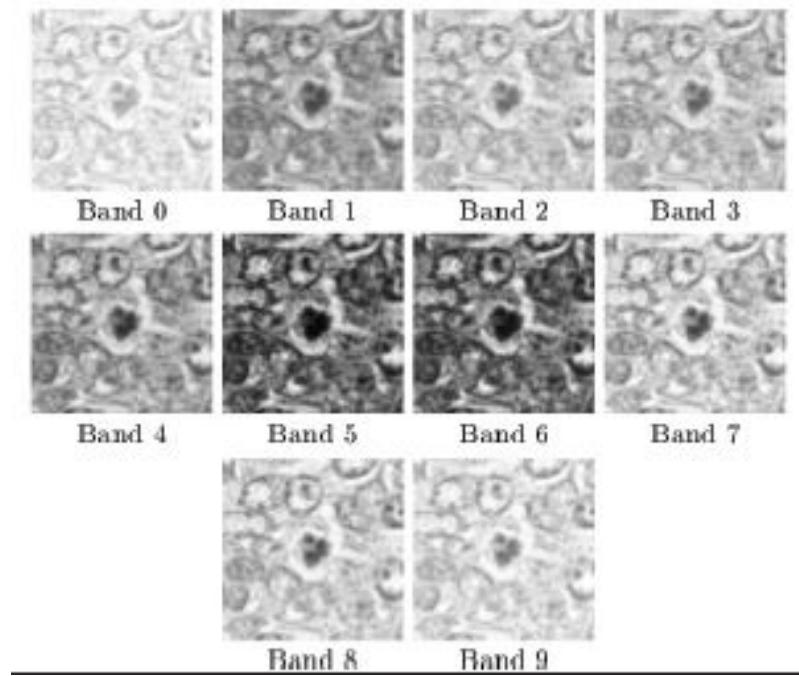


Figure 1.2 Mitotic cell on 10 different spectral bands

In the following pages of this report will be discussed related works (Chapter 2), feasibility (Chapter 3) and system analysis (Chapter 4) respectively.

2 Related works

Previous studies applied different Image Processing techniques to describe mitosis. Detailed information about all these studies can be found on the related papers. (see References)

There are plenty studies for automatic mitosis detection on RGB domain [2]–[10] using various machine learning and pattern recognition techniques[11].

In studies [5] [6] was used textural features of histopathologic images (HPI) and Local Binary Pattern to recognize MC. [4] trained Neural Network (NN) to differentiate patches with a mitotic nucleus close to the center from all other windows. [7] has taken advantage of CNN using 2 models. First model retrieve the mitosis candidates. The retrieved candidates are fed into the second model for further discrimination of mitoses and mimics with similar appearance.

On the Multi-spectral domain there are fewer studies than on normal RGB domain. [9] focused on the intensity and texture of the object in the LDA generated discriminative image space to differentiate the MCs and other cytological components. Performance of this study is F-measure 0.4790. [8] has approach based on the simultaneous consideration of spatial and spectral relationships in the detection of MC in digital multispectral HPI. For classification problem was developed different clusters using support vector machines (SVM), random forests (RF), naive Bayes (NB), and k-Nearest Neighbor classification methods. The highest F-measure gained with SVM classification (with 30-fold cross validation) is 0.5465.

According to [10] up to 129 teams have registered to the ICPR 2012 contest. However, only 17 teams submitted their detection of mitotic cells. And only 4 teams submitted multispectral image evaluation. The highest F-measure along them gained 0.5890. Detection MC on multispectral images is challenging than detection MC on images captured by scanner A and H.

3

Feasibility

In this chapter will be discussed feasibility of the thesis. The feasibility studies are given below: Technical, Legal, Economic and Time feasibility, respectively.

3.1 Technical Feasibility

There is no need for special equipment to perform the thesis except a computer with installed frameworks and programs:

- OpenCV
- Spyder
- Colab

As a programming language Python was selected because it has many open-source image processing and machine learning utilities.

3.2 Legal and Economic Feasibility

Since open-source frameworks and programs will be used there is no need for any legal or economical requirements. Project will be performed on one computer. For deep learning training task will be taken advantage of Google Colab which is remote Tensorflow-GPU powered Python Notebook.

3.3 Time Feasibility

Time feasibility Gannt diagram can be seen on Figure 3.1

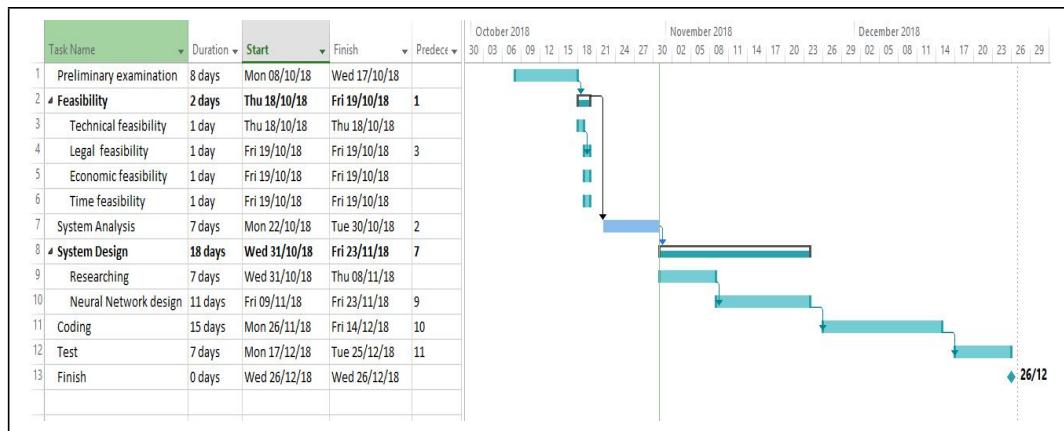


Figure 3.1 Time feasibility: Gannt diagram

4

System Analysis

In this chapter will be discussed the analysis of the system.

Mitosis detection is difficult because mitosis are small objects with a large variety of shapes, and they can thus be easily confused with some other objects or artefacts present in the image.

ICPR 2012 contest contain a total of 326 mitotic cells on images of both A and H scanners, and 322 mitotic cells on the multispectral microscope. The data set is made up of 50 high power fields (HPF) coming from 5 different slides scanned at x40 magnification. There are 10 HPFs per slide. A HPF has a size of $512\mu m \times 512\mu m$ (that is an area of $0.262 mm^2$). The pathologist has annotated all the mitotic cells manually. [10]

The camera attached on top of the multispectral microscope generates images of 1360x1360 pixels. However, to cover an area of $512\mu m \times 512\mu m$, 2767x2767 pixels are needed. Therefore, four images used to cover the same area as the A and H scanners. However, these four images do not cover completely the $512\mu m \times 512\mu m$ area, 47 pixels are missing in width and in height to cover fully the area. Each image, covering a quarter of a scanner image, is labeled a, b, c or d depending on its position in the scanner image.

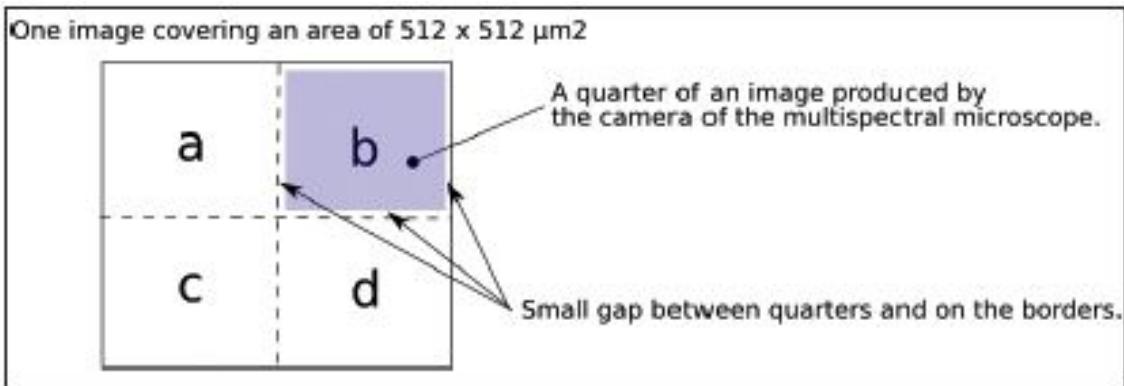


Figure 4.1 Locations of quarters on multispectral images

Figure 4.1 shows the location of each quarter a, b, c, d. As the quarters do not cover completely the $512\mu\text{m} \times 512\mu\text{m}$ area, compared to the scanner images, there is a small gap on the borders, and also a small gap between quarters a, b, c and d.

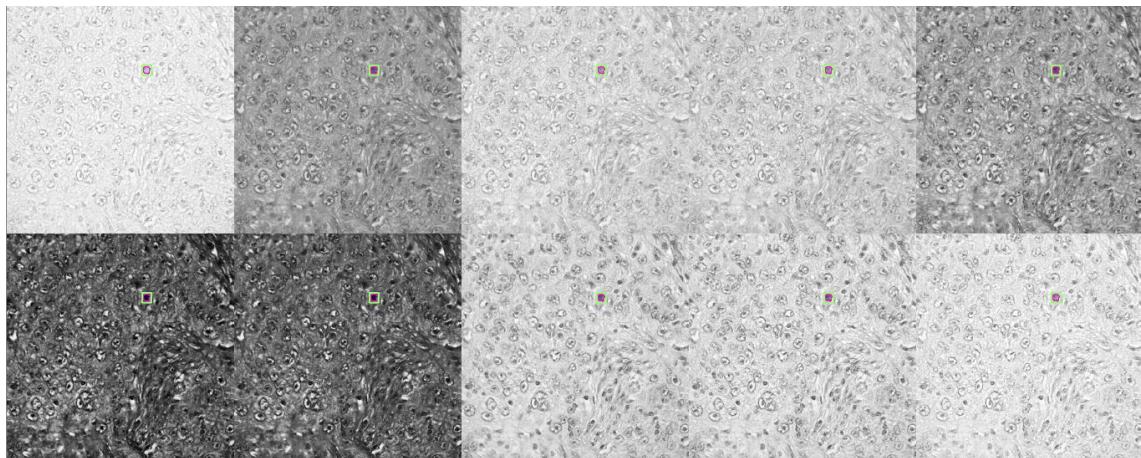


Figure 4.2 Annotated mitosis on multispectral space

Figure 4.2 shows sample multispectral image with annotated mitosis on ICPR dataset.

5

System Design

In this chapter the design of mitosis detection system will be discussed.

5.1 Image Data

The image data used in this paper are obtained from publicly available MITOS dataset[1], which includes 35 high-resolution multispectral images. All images have been manually examined by two experienced pathologists and a total of 224 MCs have been labeled. The images were acquired using a ten-band multispectral microscope.

Train set with 2 classes that are mitosis (C1) and non-mitosis (C2) generated over images. C1 class obtained with annotated MC pixels and resized to 60x60 with 10 bands. In deep learning tasks, a lot of data is needed to train DNN model. C1 class dataset is not big enough for doing DNN tasks. In order to overcome lack of data C1 class data flipped vertically, horizontally and mirrored (both horizontal and vertical flip), also for every flipped image, regions zoomed by expanding pixels with different size.

C2 class data candidates obtained from filtered Discriminative Images (DI), that is combined features of 10 different bands reduced to 1 band. DI obtained using technique proposed in [9] and explained in section 5.2 of this chapter. Candidate selection process has some steps that firstly DI applied 11x11 filter with values 1/53. This filter picks up dark pixels so that generally mitotic regions are dark pixels. Then k-means with k=2 clusters applied to separate background from candidate regions. After that regions bounded with rectangles picked and compared to true MC pixels from annotation file. If the pixels does not belong to MC, region annotated as C2 data. All C2 data resized to 60x60 and saved with 10 bands. Since lots of candidates obtained from dark regions there is no need to flip C2 data.

5.2 Implementation of Discriminative Image

As described in [9] the images with the 10 different spectral bands contain more biological signals than a single spectral image or the RGB image (three spectra), which will help to identify the mitosis. However, there is a need to find a way to combine these band spectral images and perform the MC detection. The LDA technique used to find a projection of the high-dimension data into a lower dimension space such that the best discriminant between two or more classes is achieved.

The DI is generated using the following equation:

$$I_D = AI = [a_1, \dots, a_K][I_1, \dots, I_K]^T = \sum_{i=1}^K a_i I_i \quad (5.1)$$

where a_i is the coefficient for each spectral image, K is the number of spectral bands, and I_i is a spectral image (i is the spectral index).

Using this technique requires to find a projection vector A. In order to obtain this vector firstly pixels in C1 and C2 classes are separated. Then, mean intensity (M_{C1}, M_{C2}) and variant (Σ_{C1}, Σ_{C2}) of pixels are calculated for every band.

$$M_{C1} = [\mu_{C1,1}, \mu_{C1,2}, \dots, \mu_{C1,K}] \quad (5.2)$$

$$M_{C2} = [\mu_{C2,1}, \mu_{C2,2}, \dots, \mu_{C2,K}] \quad (5.3)$$

$$\Sigma_{C1} = [\sigma_{C1,1}, \sigma_{C1,2}, \dots, \sigma_{C1,K}] \quad (5.4)$$

$$\Sigma_{C2} = [\sigma_{C2,1}, \sigma_{C2,2}, \dots, \sigma_{C2,K}] \quad (5.5)$$

Finally the projection vector A is obtained as below:

$$A = (\Sigma_{C1} + \Sigma_{C2})^{-1}(M_{C1} - M_{C2}) \quad (5.6)$$

Projection vectors has been calculated in over 100 annotated images. Average of these vectors, which is 1 vector with discriminative values will be used in generating all DI using equation 5.1

5.3 Implementation of CNN model

The CNN model architecture which used on training dataset is shown on figure 5.1. Input layer has dimension of 60x60x10 images. Conv1 layer has 96 filters with 7x7 kernel. Next Conv2 layer has 384 filters with 5x5 kernel size. Conv3 layer has 128 filters with 3x3 kernel followed by 2x2 max pooling layer. Conv4 layer has 256 filters with 3x3 kernel. And last Conv5 layer has 96 filters with 3x3 kernel followed by 2x2 max pooling layer. After convolutional layers begins fully-connected layers with 64 followed by 32 Denses. Final layer with softmax activated Dense has 2 output that is prediction of MC and non-MC data.

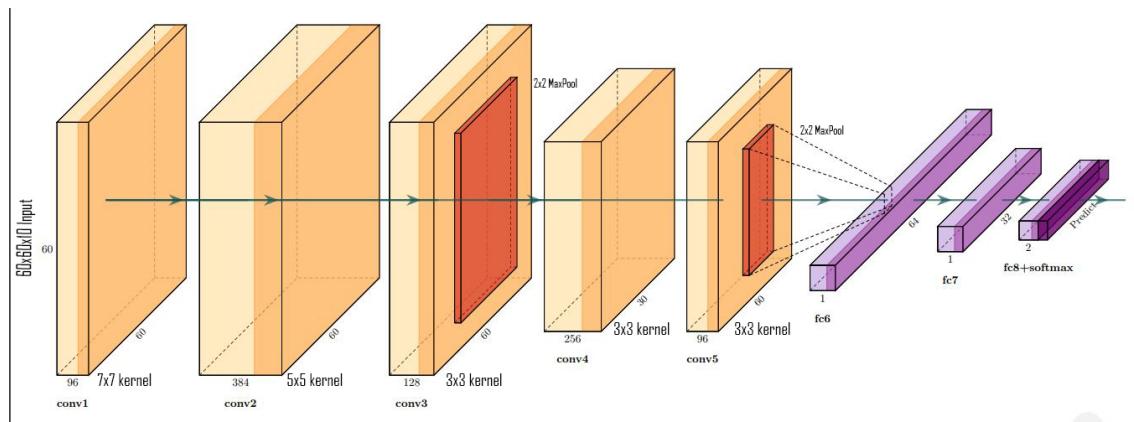


Figure 5.1 CNN model architecture

5.4 Generel overview of proposed technique

Firstly, multispectral images given as input reduced to one dimension image which called DI. Then DI is applied 2D filter. Result of filtered DI is image with only darker pixels. After distinguishing dark pixels k-means applied to separate white background and unify values of all darker pixels. Next step is applying threshold in order to get contours of ROI (Region of Interest). These ROI matched in multispectral images and resized 60x60x10 as input of CNN. CNN predicts whether given ROI mitotic or not.

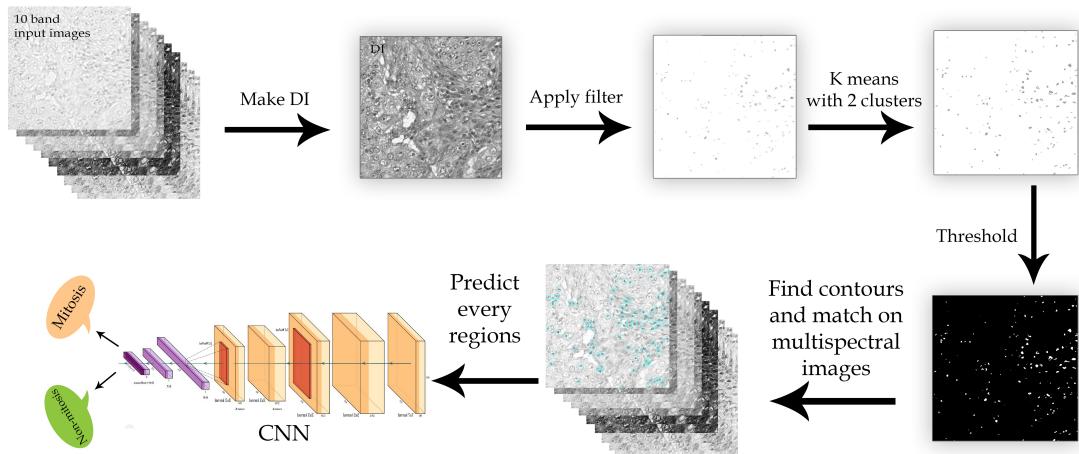


Figure 5.2 Generel steps flowchart

6

Application

In this chapter will be shown results of proposed techniques.

On Figure 6.1 shown multispectral data and generated DI image with masked mitosis region.

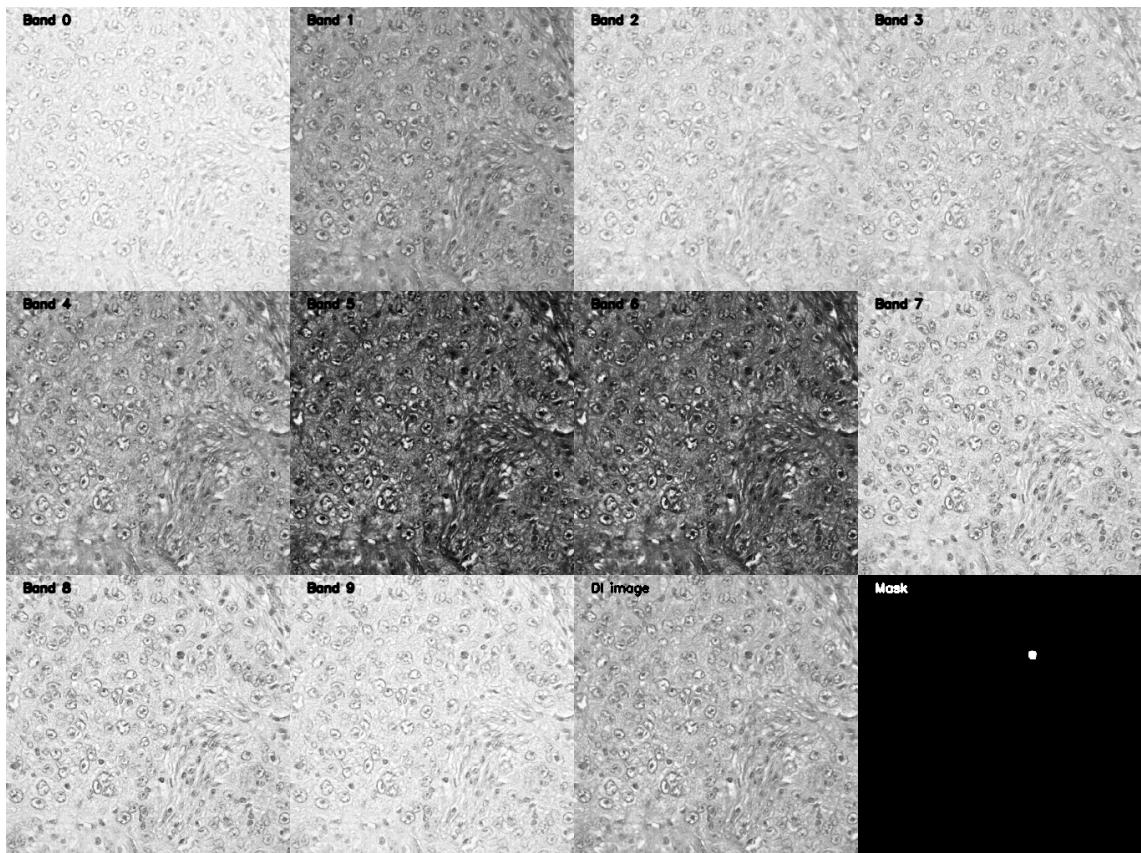


Figure 6.1 Sample image with 10 spectral bands and generated DI image

Figures 6.2 and 6.3 shows MC and non-MC region samples over 10 bands that will be used in classification problem.

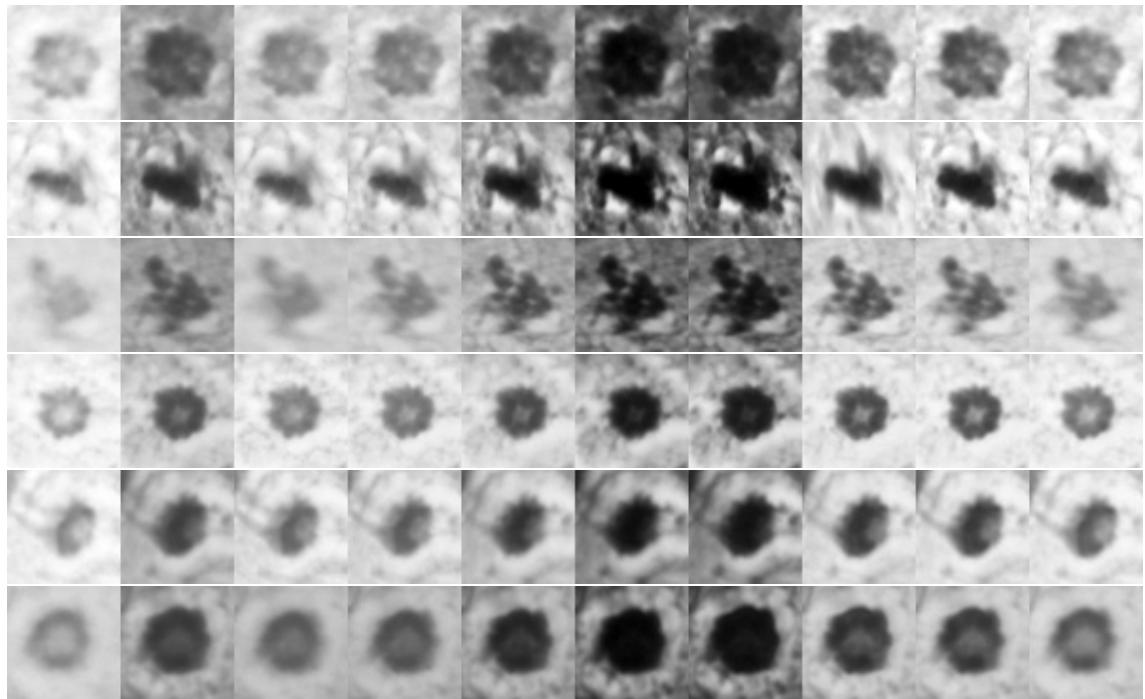


Figure 6.2 Samples for C1 class with 10 bands (mitotis regions)

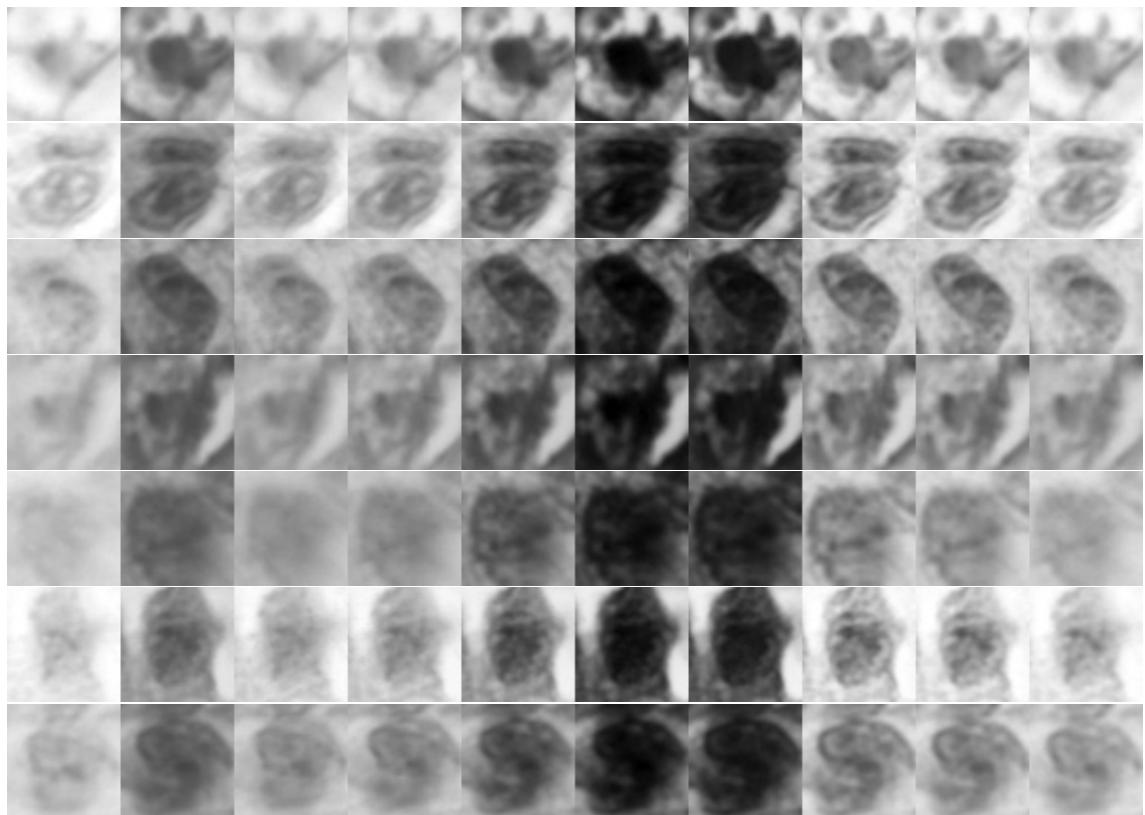


Figure 6.3 Samples for C2 class with 10 band (non-mitotis regions)

7

Experimental Results

This chapter covers results from CNN prediction.

On Figures below shown candidates for MC and actual predicted MC regions. Black circled areas are true mitoses. On Figures 7.3 and 7.4 has 1 False Positives each.

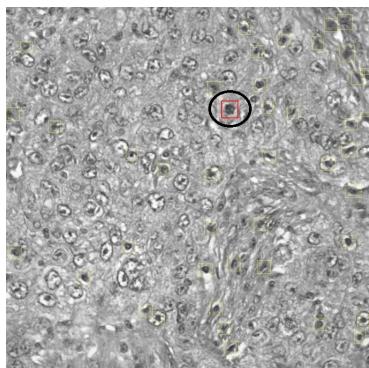


Figure 7.1 Mitosis prediction result-1

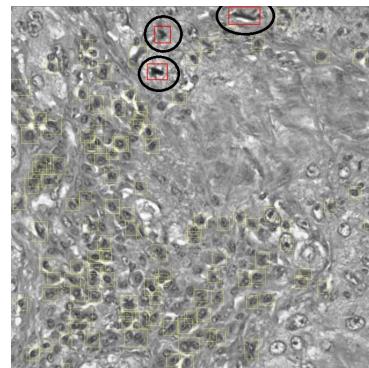


Figure 7.2 Mitosis prediction result-2

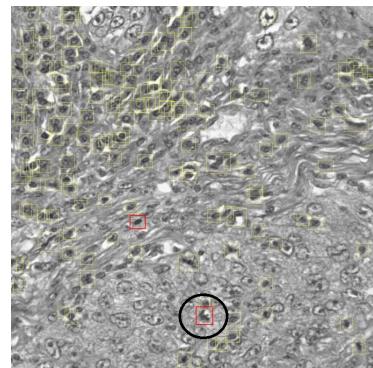


Figure 7.3 Mitosis prediction result-3

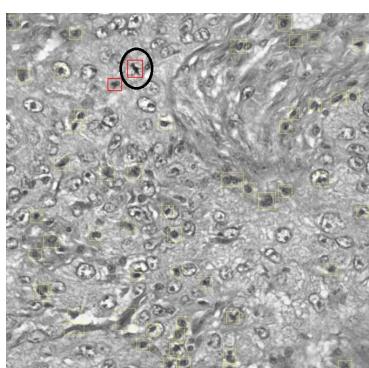


Figure 7.4 Mitosis prediction result-4

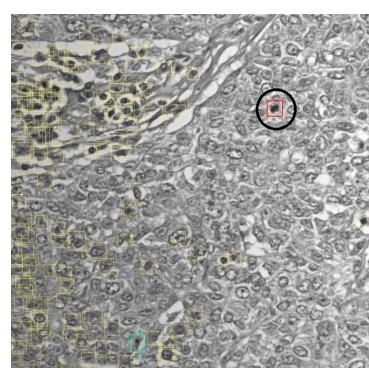


Figure 7.5 Mitosis prediction result-5

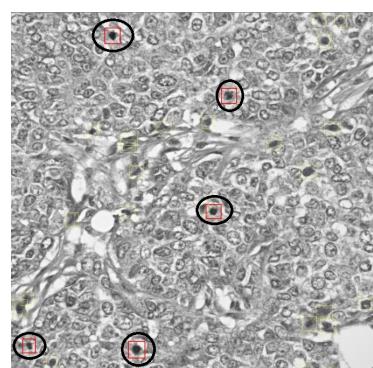


Figure 7.6 Mitosis prediction result-6

8

Performance Analysis

On this chapter performance of this work will be evaluated.

True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) rates among ICPR-2012 training dataset images is shown on the table below:

Table 8.1 Confusion table

	Predicted MC	Predicted Non-MC
Actual MC	TP=191	FN=73
Actual Non-MC	FP=47	TN=9398

$$\text{Recall (sensitivity)} = \frac{TP}{TP+FN} = \frac{191}{191+73} = 0.723$$

$$\text{Precision (positive predictive value)} = \frac{TP}{TP+FP} = \frac{191}{191+47} = 0.802$$

$$\text{F-measure} = 2 * \left(\frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \right) = 2 * \left(\frac{0.802 * 0.723}{0.802 + 0.723} \right) = 0.760$$

$$\text{Kappa} = \frac{(TP+TN)-((TP+FN)*(TP+FP)+(FP+TN)*(FN+TN))}{1-((TP+FN)*(TP+FP)+(FP+TN)*(FN+TN))} = 0.998 \text{ (due to high TN rate)}$$

9 Conclusion

In this work, a deep learning based feature extraction method using Convolutional Neural Networks is proposed for automated mitosis detection on histopathological images. The proposed method has been tested on mitosis detection in breast cancer histopathological images (ICPR-2012) data set. A direct classification approach can not be accomplished accurately because of the imbalanced number of mitotic and non-mitotic cells. Mitotic cells training data were zoomed by expanding pixels around candidate regions and also flipping them horizontally, vertically and both horizontal-vertically. And non-mitotic cells were acquired from dark regions on Discriminative Image. Approximately 50,000 non-mitotic and 14,000 mitotic regions has been generated from original 10-band images. The precision, recall and F-measure values are 0.802, 0.723, 0.760 respectively. These results prove that the proposed technique achieved promising results for mitosis detection on histopathological images. Evaluation on the public available dataset and comparison with the existing technique show the effectiveness of the proposed technique. The proposed technique is expected to reduce the workload of pathologists when they evaluate the cancer grade of biopsy.

References

- [1] (2012). Mitosis detection in breast cancer histological images, [Online]. Available: <http://www.icpr2012.org/>.
- [2] C. Sommer, L. Fiaschi, F. A. Hamprecht, and D. W. Gerlich, “Learning-based mitotic cell detection in histopathological images,”
- [3] B. Wu, B. Fan, Q. Xiao, T. Kausar, and W. Wang, “Multi-scale deep neural network for mitosis detection in breast cancer histological images,”
- [4] H. Chen, Q. Dou, X. Wang, J. Qin, and P-A. Heng, “Mitosis detection in breast cancer histology images via deep cascaded networks,”
- [5] A. Albayrak and G. Bilgin, “Detection of mitotic cells in histopathological images using textural features,”
- [6] İ. O. Sığircı, A. Albayrak, and G. Bilgin, “Detection of mitotic cells using completed local binary pattern in histopathological images,”
- [7] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Mitosis detection in breast cancer histology images with deep neural networks,”
- [8] H. Çukur and G. Bilgin, “Detection of mitotic cells in multispectral histopathological images,”
- [9] C. Lu and M. Mandal, “Toward automatic mitotic cell detection and segmentation in multispectral histopathological images,” *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS*, vol. 18, no. 2, 2014.
- [10] L. Roux, D. Racoceanu, N. Loménie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. L. Naour, and M. N. Gurcan, “Mitosis detection in breast cancer histological images an icpr 2012 contest,”
- [11] R. Polikar, “Pattern recognition,”

Curriculum Vitae

FIRST MEMBER

Name-Surname: Sardor HAZRATOV

Birthdate and Place of Birth: 04.02.1995, Türkmenistan

E-mail: serdarjan1995@gmail.com

Phone: 0507 037 75 38

Practical Training: ENUYGUN İNTERNET BİLGİ HİZM.A.Ş.
KAFEİN YAZILIM HİZM.A.Ş.

Project System Informations

System and Software: Windows, Anaconda, Spyder, Python, Google Colab.

Required RAM: 8GB

Required Disk: 1TB